The Price of Caution near a Collapsing Threshold

Romain Fillon[1,2]

[1]*Paris-Saclay Applied Economics, Graduate School of Economics & Management, Université Paris-Saclay, France*
[2]*CIRED, Université Paris-Saclay, France*

August 22, 2026

Abstract

Standard risk theory says that uncertainty raises the value of caution. It puts the uncertainty in where the threshold lies. Here it lies in the system itself: the threshold is a fixed feature of a state that fluctuates, and the fluctuations are what carry the system across it, so what the decision-maker controls narrows the distance they must cover. Noise and control are therefore substitutes, and the sign reverses: lower the noise in a given system and its price of caution rises rather than falls. How sharply that price rises as the threshold nears is fixed by the rate at which the distance closes only, not by the kind of threshold or the loss a crossing would cause. Climate satisfies both conditions: its elements fluctuate enough to cross unaided, and every emissions scenario now under consideration narrows the distance they must cross. The framework sorts climate elements in two: the price of caution grows without bound for the Atlantic overturning circulation, the Amazon and the West Antarctic ice sheet (WAIS), and stays bounded for oscillatory regimes such as El Niño. The price also rises with the discount rate where the conventional carbon price falls, and a backstop caps it at a date that is a decarbonization deadline, first for the WAIS. What a risk costs society has been treated as a question about the event: how likely, how bad, how soon. Much of the answer turns out to lie in the geometry of the approach: not the system's own dynamics, but the rate at which spending the budget closes the distance to its edge.

1 Introduction

Economics has a settled answer to what uncertainty does to the value of caution, and the answer is that it raises it. Option values, precautionary saving and risk premia all grow with the variance of the shocks a decision-maker faces, a logic that runs from Pindyck (2000) to the risk-adjusted carbon prices of van den Bremer and van der Ploeg (2021). That answer rests on a loss which is a smooth convex function of the state, so that a wider distribution carries a heavier expectation. This paper studies the case in which the loss is triggered instead by crossing a boundary, and in which the decision-maker moves the drift of the state rather than his exposure to it. Under those two features the sign reverses. Hold the distance to the edge fixed, lower the noise, and the marginal value of not taking one more step toward the boundary goes up. A calmer system commands a higher price of restraint.

What a marginal unit of restraint buys is a later crossing and not a safer one: the boundary is reached with probability one, and two forces set the date at which it is reached, the noise in the state and the steady push that spending a controllable budget adds. The two are substitutes in bringing that date forward, so each is worth less at the margin when the other is abundant. When the noise is large the system goes over the edge on noise alone, whatever the planner does at the margin. When the noise is weak it rarely suffices on its own, and it is the planner's own push that sets the pace, so each unit of restraint bears directly on when crossing occurs. Restraint is worth most when it is the only thing taking the system to the edge. The sign is a theorem and not a numerical finding: it follows from a single monotonicity property of the crossing profile, and it holds at every discount rate as long as the noise remain strong enough that the threshold is what decides the outcome. Nothing in that argument mentions carbon. It uses only the effect of a certain increment on a first-passage probability, so it applies to any controllable stock drifting toward a collapsing boundary, whether the stock is carbon, leverage, or a harvested population. For instance, it gives a reading of the volatility paradox of Brunnermeier and Sannikov (2014), in which low exogenous volatility encourages leverage accumulation, raising endogenous systemic risk, so that endogenous leverage rather than exogenous risk sets the distance to crisis. Our mechanism reaches the same conclusion from the marginal value of restraint: the calm is not incidental to the buildup, but what makes each further increment in the controlled drift matter.

Because the reversal holds across threshold geometries, it is silent on what a planner most needs to know, which is how violently the price rises as the edge approaches. That is settled by a single exponent. Write ϵ for the fraction of the budget to the threshold still unspent, so that ϵ falls from one toward zero as the system is pushed toward the edge. As the budget runs out, the stable state and the boundary move toward each other, and the safety margin between them closes at a definite rate. A margin closing as a square root, which is the fold of a saddle-node, makes the price diverge as $\epsilon^{-1/2}$. A margin closing linearly leaves it bounded. What decides divergence is that closing rate, not the name of the bifurcation, not the size of the noise, and not the damage at crossing. The closing rate is a property of how the control couples to the threshold rather than of the phase portrait on its own, and the distinction matters: two geometries, the fold and the transcritical, have the same phase portrait, one being a translation of the other, and they fall in opposite classes because the budget reaches their margins at different rates. The literature on climate tipping elements has been organized around two axes, the damage in climate economics and the location of the warming threshold in climate science. Neither decides how sharply the price rises as the threshold nears. It is decided by how fast the margin closes as the budget runs out.

A threshold can collapse in more than one way, and the way it does is what decides the class. Neither the reversal nor the strength of the noise effect can say which one an element belongs to: the first holds in every geometry, and as for the second, two thresholds can have identical dynamics as they are approached, and therefore the same leading elasticity, while their margins close at different rates and they fall on opposite sides. What separates them appears one order further out, in the way the noise effect itself changes as an element approaches its edge. Compare two elements at equal distance to threshold, one quietly driven and one noisily driven, and follow the ratio of their prices as the budget is spent down. That ratio widens with the remaining budget in both classes, but faster in the divergent one, by a factor running from a few units at the far end of the range we can compute to two orders of magnitude close to the threshold, so the separation between the two classes deepens as the edge is approached. Closer to the threshold the contrast changes character. Well beyond where the calibrated elements now sit, the two classes stop differing in the size of the effect and start differing in its exponent: the price ratio moves as the square root of the remaining budget in the divergent class, and as the budget itself in the bounded one. That regime is out of reach of the instrumental record for climate elements, since no record supplies a proximity that close, but within reach of a designed ensemble of simulations.

At any fixed budget the boundary sits a finite distance away and the noise eventually carries the state across it, so crossing the threshold occurs in finite time almost surely, though that time can run to millennia for an element far from its edge. What restraint buys is therefore time rather than the certainty of never crossing. Postponement is worth something only to a planner who discounts, and it is worth more to a planner who discounts more. The threshold component of the carbon price consequently rises with the discount rate, where the ordinary Ramsey component falls. The strength of that reversal depends on how fast the damage arrives: impatience raises the value of delay and at the same time discounts a loss that unfolds slowly, so over the range that separates Nordhaus from Stern the threshold price roughly doubles for an element realizing its damage in decades and is flat for one realizing it over millennia. The unresolved disagreement about discounting therefore moves the two components of the price in opposite directions, by amounts that depend on the element.

Climate change is where the classification has consequences, and its elements do not all fall on the same side. The Atlantic overturning circulation, the Amazon, and the West Antarctic and Greenland ice sheets are folds, so their tipping price diverges as the budget to their thresholds runs out. Expected-damage reasoning smooths over that singularity: it represents crossing by a rate that stays finite as the state approaches its edge, and a finite rate delivers a finite price. Finiteness is what the fold denies, the margin the noise must cover closing with a vertical tangent, so the last tonne destroys unboundedly more of it than the first. Oscillatory regimes such as the El Niño–Southern Oscillation are bounded and are priced as ordinary externalities. The assignment requires only the rate at which the margin closes, which the structural climate literature reports independently of any damage function (Lenton et al., 2008; Armstrong McKay et al., 2022; Ashwin et al., 2012), so a planner can act on the classification long before the damage debate is settled.

A divergent price would be of little use if it were merely a singularity, and this one is not, because the optimal path never reaches it. Pairing the price with a backstop available at a finite cost produces an endogenous budget floor and a switching date at which decarbonization becomes optimal outright. The singularity thus turns into a deadline, and the deadlines can be ranked: the West Antarctic ice sheet comes first, with a median of about 27 years, and holds that position in 90% of draws, while the Amazon and the overturning circulation sit more than a century further out and the calibration does not separate the two of them. This is the sense in which our divergence differs from the one economists already know. This is the sense in which our divergence differs from the one economists already know. The dismal theorem of Weitzman (2009) builds an unbounded price out of an unbounded loss, a fat tail in climate sensitivity meeting a marginal utility that explodes as consumption falls. Ours holds the loss finite and lets the margin do the work: what grows without bound is the value of one more unit of restraint, because each unit destroys more of the margin than the last.

Where the calibrated elements sit relative to the divergence is a separate question. An element observed stable for many multiples of its own recovery time cannot be close to its boundary, since it would have crossed, and that argument places the overturning circulation and the Amazon outside the range in which the leading-order law holds. What the paper prices is therefore the regime these elements will enter rather than the one they occupy, and what it dates is the point within that regime at which full decarbonization becomes optimal.

Three ways of representing a threshold are current in the climate-economics literature, and ours is a fourth. The first posits the hazard as a smooth function of a climate state: Cai and Lontzek (2019) make the annual survival probability linear in temperature above a floor, and van der Ploeg and de Zeeuw (2019) make it linear or quadratic in the carbon stock. The second keeps a hard threshold but places it at an unknown location and lets the planner learn where it is (Lemoine and Traeger, 2014; Fillon et al., 2023), so that the randomness governing the crossing is epistemic rather than dynamic. The third also posits a hazard and adds that it is small: Van Den Bremer et al. (2026) treat the risk of tipping as a small parameter alongside the damage share, and expand around its absence to obtain rules whose components are separately interpretable. What the three share is a hazard that is finite wherever the state is finite, and nothing in them that collapses as the budget is spent. We keep a hard threshold, place it in the element's own state rather than in temperature, and let the noise in that state do the crossing. That last choice is what removes avoidance from the menu. In all three representations a policy exists that never crosses, whether by holding the stock short of the threshold or by returning it below the floor where the hazard vanishes, so caution buys an outcome. Here the state fluctuates and the boundary lies a finite distance away, so it is reached in finite time with probability one wherever the budget stops; ending emissions freezes the margin, and a frozen margin makes the expected wait exponentially long rather than infinite. Caution buys a date, which is what inverts the comparative static in the discount rate below. Our closest relative by machinery is van den Bremer and van der Ploeg (2021), who price climate risk in a Pindyck and Wang (2013) growth model under recursive Epstein–Zin preferences, by a regular expansion in the damage ratio, and obtain a positive risk premium with no threshold anywhere in the model. We take that machinery to the threshold they set aside, where the expansion turns singular and the sign inverts.

The framework does not deliver a figure in dollars. The model locates an element in the units of its own equation of motion, while a climate record measures the same distance in the units of an instrument, degrees for the overturning circulation and a vegetation index for the Amazon. Converting one into the other takes one number per element, the margin between the current stable state and the unstable one expressed in the element's own observable, and no element has it. Everything the paper claims survives that gap, because every claim is a proportion rather than a level: the divergence exponent, the sign of the noise effect and the cross-partial that separates the two classes are ratios in which a conversion factor cancels. A price per tonne is a level, and it survives only up to that number. For the overturning circulation, the one element whose recovery rate and variability are both on record, we report it as a curve in the margin measured in standard deviations of that record, rather than as a point estimate. Naming that number is also where this paper meets a step the literature has begun to take, which is to model the crossing as a property of an element's own dynamics rather than as a posited probability. Fillon and Guivarch (2023) replace that probability by a calibrated vegetation system for the Amazon and price the subsystem as a state variable, and in such a system the conversion problem does not arise: the state is already a share, so the margin and the variability of forest cover are in the same units and their ratio is what the price needs. That system also carries both of our classes at once, a fold at the dieback and a transcritical at lower stress. What is added here is the law those dynamics obey near the edge, since the exponent, the sign of the noise effect and the classification follow from the rate at which the margin closes and from nothing else.

Section 2 develops the general result, Section 3 reads them as a classification of climate tipping elements, and Section 4 calibrates the result and turns it into policy deadlines.

2 How geometry guides the price

A planner faces a permanent welfare loss $\mathcal{L}$ at the instant a state X_t crosses a boundary that it has been keeping at a distance. Spending down a controllable budget pushes the state toward that boundary; the object we price is the marginal value of restraint, the shadow price $\Pi = -\partial V/\partial M$ of one more unit of budget spent now, where V is the planner's value function and M the accumulated stock. Climate is the leading case: X a climate indicator, the budget the remaining carbon budget, Π the threshold component of the social cost of carbon. But nothing in what follows is specific to carbon.

Two results govern this price. First, when the planner controls the system's drift, i.e. its steady push toward the edge, a quieter system commands a higher price of restraint: lowering the noise raises Π, the opposite of the intuition that more uncertainty raises the price of risk. Second, whether Π diverges or stays bounded as the edge nears is fixed by the closing rate of the safety margin, i.e. how fast the safety margin closes as the budget runs out, either square-root (divergent) or linear (bounded), not by the type of system.

Because the loss is triggered only by crossing, and crossing occurs with probability one, what a marginal unit of restraint buys is postponement: the loss arrives later and is discounted more heavily. Two forces set the date, the noise in the state and the deterministic push that spending the budget adds. When noise is large, the system crosses on noise alone whatever the planner does at the margin, so one more unit of restraint barely moves the date and is worth little. When noise is small, the noise rarely suffices on its own and it is the steady push that carries the system over the edge; each unit of restraint then bears directly on when crossing occurs, and is worth a lot. Lowering the noise therefore raises the value of restraint.

2.1 Model

The model has three ingredients. First, a state X_t evolves under a controllable drift against noise,

$$dX_t \;=\; f(X_t;\mu)\,dt \;+\; \sigma\,dW_t, \tag{1}$$

where $\mu = M^* - M = M^*\epsilon \geq 0$ is a controllable budget: the control accumulates a stock M and shrinks μ toward 0. Second, for each $\mu > 0$ the drift has a stable rest point $x^*(\mu)$ and an adjacent unstable point $x_-(\mu)$, the boundary, which absorbs the state. The gap between them is the safety margin $d(\mu) = x^*(\mu) - x_-(\mu) > 0,\ d(\mu) \downarrow 0$ as $\mu \downarrow 0$, which closes as the budget runs out. Third, crossing the boundary at the random time $\tau = \inf\{t : X_t \leq x_-(\mu_t)\}$ inflicts a one-time welfare loss $\mathcal{L}$, and the planner discounts the future at rate ρ. These are the primitives: a controllable drift toward an absorbing boundary, a margin that the budget draws down, and a loss. The object we price is Π, the marginal value of restraint, which is the shadow price of spending one more unit of budget.

The loss strikes once, at the random date τ, so its value today is the loss discounted from that date and averaged over when it falls, $\mathcal{L}\,\mathbb{E}_x[e^{-\rho\tau}] \equiv \mathcal{L}\,\varphi$, with $\varphi(x;\mu) = \mathbb{E}_x[e^{-\rho\tau}]$. Thus φ is the present value of one unit of loss delivered at the crossing instant, the price today of a claim that pays one unit on crossing. It equals 1 at the boundary, where $\tau = 0$, and falls toward 0 as the stable state sits far from the boundary, since crossing is then pushed into the distant, heavily discounted future. Turning this object into a closed form carbon price takes three modelling assumptions on functional form.

The first assumption is that preferences are homothetic. Optimal consumption is then a fixed proportion of wealth, so the consumption and saving decision no longer responds to the climate state, and the spending problem separates from the crossing problem. Under non-homothetic preferences, such as a subsistence floor below which consumption cannot fall, consumption stops being proportional to wealth and the separation fails.

The second assumption is that this homothetic preference is logarithmic. Log utility makes the separation additive: the value function splits as $V = \frac{1}{\rho}\log K + v(M, X) + C_0$, so the wealth term stands apart and the climate problem is carried entirely by the remainder v. A power utility would multiply the two instead, so that the level of wealth would scale the threshold term and reintroduce a cross-derivative between capital and the budget. Log is the homothetic case in which the wealth effect is additive and the climate term stands alone.

The third assumption is that technology is linear in capital (AK). The consumption first-order condition then gives $c^* = \rho K$, a fixed share of capital independent of the climate state and the budget, and every term in K cancels from the value function, leaving a reduced problem for $v(M, X)$ alone. Without linearity the saving rate would vary with the level of capital and couple back to the climate state, and the separation would again fail. Log utility with AK technology is the growth structure of Pindyck and Wang (2013) specialised: their framework is power utility with linear technology, and log is the case of it in which the separation becomes additive rather than multiplicative.

A few lines show what the three assumptions buy. Capital accumulates as $\dot{K} = AK - c - C(a)K$ with $A > \rho > 0$, where $C(a)$ is the abatement cost as a fraction of output, and spending the budget moves the stock at rate $\dot{M} = (1-a)E_0$. The stock is carried absolutely throughout, so that $\mu = M^* - M$ and $\epsilon = \mu/M^*$ is the fraction; the factors of M^* then appear where the fraction is taken and nowhere else. The planner's Hamilton–Jacobi–Bellman equation is

$$\rho V = \max_{c,\, a\in[0,1]} \Big\{ \log c + V_K\big(AK - c - C(a)K\big) + V_M\,(1-a)E_0 + f(X;M)\,V_X + \frac{\sigma^2}{2} V_{XX} \Big\}.$$

Posit the additive form $V(K,M,X) = \frac{1}{\rho}\log K + v(M,X) + C_0$, which gives $V_K = 1/(\rho K)$ while V_M, V_X and V_{XX} are simply v_M, v_X and v_{XX}. The consumption first-order condition reads $1/c^* = V_K$, so $c^* = \rho K$: optimal consumption is a fixed share of capital, independent of the climate state and of the budget. Substituting the posited form and c^* into the equation, the left-hand side is $\log K + \rho v + \rho C_0$ and the right-hand side is

$$\log\rho + \log K + \frac{A - \rho - C(a)}{\rho} + v_M\,(1-a)E_0 + f\,v_X + \frac{\sigma^2}{2} v_{XX}.$$

The capital stock has left the problem. The term $\log K$ appears on both sides and cancels, and the return on capital survives only as a constant, because dividing the linear return AK by the marginal utility $1/(\rho K)$ removes the stock. Fixing C_0 to absorb the constants that remain and maximising over a leaves a reduced problem for the climate block alone,

$$\frac{\sigma^2}{2} v_{XX} + f(X;M)\,v_X - \rho v \;=\; -\max_{a\in[0,1]} \left\{ -\frac{C(a)}{\rho} + v_M\,(1-a)E_0 \right\}, \tag{2}$$

with the crossing penalty entering only through the boundary condition $v(M, x_-(\mu)) = -\mathcal{L}$. Log utility is what makes the split additive rather than multiplicative; the AK technology is what makes the consumption share independent of the state. Drop either and the capital stock reappears inside the climate block. The functional forms are conveniences chosen to deliver a closed form, and they do not drive the result.

With these three assumptions in place, the climate state enters welfare only through v, and within v only through the crossing penalty, whose sole trace is the boundary condition $v(M, x_-(\mu)) = -\mathcal{L}$. Spending the budget shrinks μ, moves the boundary, and raises φ, so the budget touches the climate part of welfare through φ and nothing else. Spending also shifts ordinary consumption and saving, the Ramsey channel of any growth model, but that channel lives in the capital stock alone. The two channels therefore add rather than interact, because neither depends on the other's variable, and differentiating welfare in the budget leaves no cross-term. The marginal value of restraint is the rate at which one more unit of budget raises the present-value threshold bill, so $\Pi = \mathcal{L}\,|\partial_M \varphi|$. Rewriting this in the remaining-budget fraction $\epsilon = \mu/M^*$, with $\mu = M^* - M$, gives

$$\Pi(\epsilon,\sigma,\rho) \;=\; \frac{\mathcal{L}}{M^*} \left| \tfrac{\partial}{\partial \epsilon}\,\varphi \right|. \tag{3}$$

In the climate reading, Π becomes $\mathrm{SCC}_{\mathrm{threshold}}$, the threshold component of the social cost of carbon. The price therefore reduces to one analytic question: how sensitive is the crossing value φ to the budget? The equation that answers it comes from a one-step argument. Hold the budget fixed and start the state at x. Over a short interval dt, either the state crosses and the claim pays, or it does not, in which case the planner holds the very same claim an instant later, started from the new position and discounted by one instant. So $\varphi(x) = \mathbb{E}_x\big[e^{-\rho\, dt}\, \varphi(X_{dt})\big]$: the claim price, once discounted, does not drift. Expanding both factors to first order in dt, with $\mathbb{E}[dX] = f\, dt$ and $\mathbb{E}[(dX)^2] = \sigma^2\, dt$ read off (1),

$$\varphi = (1 - \rho\, dt)\Big(\varphi + f\varphi'\, dt + \tfrac{\sigma^2}{2}\varphi''\, dt\Big) + o(dt),$$

and cancelling φ and dividing by dt leaves

$$\underbrace{\tfrac{\sigma^2}{2}\,\varphi''}_{\text{diffusion}} + \underbrace{f\,\varphi'}_{\text{drift}} = \underbrace{\rho\,\varphi}_{\text{discounting}}\,, \qquad x \in (x_-, \infty). \tag{4}$$

The two boundary conditions are read straight off the definition: at x_- crossing is immediate, $\tau = 0$, so $\varphi(x_-) = 1$; far from the boundary crossing is pushed into an ever more distant future, so $\varphi(\infty) = 0$. This is the Feynman–Kac representation of the discounted crossing claim, and the budget enters through the drift: spending it lowers μ, which brings the unstable state toward the stable one and shortens the distance φ has to discount over[1].

We now study how Π behaves as the boundary nears, on the range of proximities where the budget barely moves during a crossing, and as the noise σ changes.

2.2 Three assumptions

Whether the theorem applies to a given system depends on three properties of its dynamics. They rest on two quantities, which we define first.

The first quantity is the closing rate $p > 0$, paired with a constant $c > 0$. It records how fast the safety margin closes as the budget runs out,

$$d(\mu) \;=\; x^*(\mu) - x_-(\mu), \qquad \lim_{\mu \downarrow 0} \frac{d(\mu)}{c\,\mu^p} = 1,$$

so that d behaves as $c\mu^p$ near the threshold. One reads p off the way the stable and crossing equilibria of $f(\cdot;\mu)$ merge as the budget vanishes. A small p means the margin collapses fast, a large p means it closes gently.

The second quantity is the noise-layer width $\ell(\sigma)$, the width of the narrow zone around the threshold in which random noise and the deterministic restoring force are of comparable

[1] We compute φ in an approximation that holds μ fixed while the state crosses. Strictly φ depends on both, and the exact equation is a partial differential equation carrying an additional term $\dot\mu\, \partial_\mu \varphi$; (4) is an ordinary differential equation because μ is treated as a parameter rather than a state. It is the leading term when the budget is spent down slowly relative to the system's own recovery.

strength. Outside that zone one of the two dominates and the outcome is settled; inside it they are comparable, which is where a marginal push can still change what happens. Four steps fix the size of $\ell(\sigma)$. The first reads the drift at the threshold, since that is what the noise has to compete against. The second asks how large each term of the pricing equation would be if the crossing value varied on a length ℓ. The third shows that two of those terms must be of comparable size, and the fourth solves that condition for ℓ.

The first step is to read the drift at the threshold. Setting $\mu = 0$, where the two equilibria have merged, and measuring the state from that merged point gives the drift $f(x;0)$. Expanding it about the origin $f(x;0) = f(0;0) + f'(0;0)\,x + \frac{1}{2}f''(0;0)\,x^2 + \ldots$, the first two terms vanish, and for different reasons. The constant vanishes because the origin is an equilibrium, where the drift is zero by definition. The linear term vanishes because two equilibria have just merged there: while they were distinct the drift crossed zero between them with a definite slope at each, and once merged it only touches zero rather than crossing, which leaves it flat. Something must survive, since a drift whose derivatives all vanish is identically zero and has no dynamics at all. We call k the index of the first surviving term and write its coefficient $-c_k$ with $c_k > 0$, so that

$$f(x;0) = -c_k\,x^k\big(1 + O(x)\big), \qquad x \to 0. \tag{5}$$

This is why a noise layer exists at all, however small the noise. The restoring force does not merely weaken near the merged point, it vanishes there, so however small σ is, there is always a neighbourhood close enough to the threshold that the noise wins. Lowering the noise does not remove the zone in which the outcome is contested; it only shrinks it. That is the mechanism behind the reversal of Section 2.4: a narrower zone means the crossing value falls more steeply as the state retreats from the edge, and a steeper fall means a larger price for one more step toward it.

The second step is to size the three terms of (4) when φ varies on a length ℓ. The crossing value equals one at the boundary and falls away from it, so it changes by an amount of order one across the layer. A first derivative is a change divided by the length over which it occurs, so $\varphi' \sim 1/\ell$, and each further derivative costs another factor of $1/\ell$, so $\varphi'' \sim 1/\ell^2$. The diffusion term is σ^2 times that second derivative, hence of order σ^2/ℓ^2. The drift term takes one extra step, because its coefficient is not constant: evaluating (5) at a distance ℓ from the merged point, which is where the layer reaches, gives $|f| \sim c_k\ell^k$, and multiplying by $\varphi' \sim 1/\ell$ leaves ℓ^{k-1}. The discount term is of order ρ and carries no power of ℓ at all, so it says nothing about which length the crossing value varies on and cannot be one of the two that fix it. The balance is therefore between the two that remain.

The third step is to see that diffusion and drift must balance, which follows by eliminating the alternatives. Were the layer much wider, the diffusion term would be negligible and the equation would collapse to $f\varphi' = \rho\varphi$. That equation is first order, so it admits one constant of integration and cannot meet both $\varphi(x_-) = 1$ and $\varphi(\infty) = 0$; more tellingly, the noise has left it, and it describes the deterministic problem in which a state resting at a stable equilibrium never reaches the boundary and the crossing claim is worthless. Were the layer much narrower, the drift term would be negligible and the equation would collapse to $\frac{\sigma^2}{2}\varphi'' = \rho\varphi$. That one keeps the noise but loses the geometry, since k has disappeared from it and a system with $k = 2$ could not be told from a system with $k = 3$. Escape near a threshold is precisely the contest between the pull back and the noise, so the only length at which the equation still describes it is the one at which neither term dominates.

The fourth step is arithmetic. Setting the diffusion term and the drift term of the previous step equal to each other, which is what defines the layer $\sigma^2/\ell^2 \sim \ell^{k-1} \leftrightarrow \ell(\sigma) = \sigma^{2/(k+1)}$.

Two points about that balance are worth adding before the hypotheses. The first concerns why the width is read at $\mu = 0$. It is a property of the limiting problem at the threshold, and at that point the terms of the drift that carry the budget have vanished by construction. Away from the threshold they return, and two lengths are then in play. The margin $d(\mu)$ is the distance the state must cover to reach the boundary, and the budget sets it; the layer $\ell(\sigma)$ is the width over which the crossing value varies, and the noise sets it. The restored terms vary over the first and leave the second untouched, which is why ℓ may be read at $\mu = 0$ without loss. The price keeps them apart in the same way: the margin enters through the proximity, the layer through the steepness of the crossing profile. The second concerns the term the balance left out, because it settles how impatience enters. Of the three terms sized above (diffusion, drift, discounting), the discount term was the only one carrying no power of ℓ. Dividing the equation by the common size of the other two therefore divides the discount rate along with everything else, so what survives is not that rate but its ratio to a power of the noise. A planner's impatience reaches the near-threshold problem only when measured against the scale of that problem, which is what the second hypothesis below asserts.

The three hypotheses follow. The first, (H1), requires that the margin close at a definite rate, so that p and c exist and are positive. As stated it is a limit, and a limit constrains d without constraining d': two margins can share an asymptote while their derivatives disagree. The price needs the derivative, since it differentiates the crossing value with respect to the budget, so the asymptotic relation must be one that survives differentiation. It does here. The equilibria are roots of a smooth function, so the implicit function theorem gives d the regularity it needs and $d'(\mu) \sim c\,p\,\mu^{p-1}$ follows; Table 1 carries the four geometries, for each of which the constant and the exponent are read off in closed form.

The second, (H2), names the ratio that the previous paragraph left implicit. Near the threshold the noise level and the discount rate stop acting separately: what matters is not how impatient the planner is, but how impatient relative to the speed at which the system crosses the layer $\ell(\sigma)$. The hypothesis is that nothing else survives: neither the noise level nor the discount rate enters the price on its own, only the ratio $\tilde{\rho} = \rho/m(\sigma)$, where $m(\sigma) = \sigma^b$ is that crossing speed. That second scale is determined rather than assumed: rescaling the equation leaves $b = 2(k-1)/(k+1)$, positive for every $k \geq 2$, so the requirement follows from two equilibria merging rather than restricting anything further (Appendix A.2).

The third, (H3), concerns the one number the price reads off that rescaled problem. That number is a steepness: how fast the value of a claim paying at crossing falls as the state is moved back from the boundary, measured in widths of the layer. We call it the amplitude $\Phi(\tilde{\rho})$, and it is a slope rather than a level because the rescaled problem fixes its solution only up to a multiplicative constant, so anything read off it has to be free of that constant. A large amplitude describes a system in which one more step toward the boundary buys a large increase in the present value of crossing. What the hypothesis requires is that this steepness be positive and increasing in $\tilde{\rho}$, so that a more impatient planner faces a value falling away more sharply from the edge. Appendix A.3 establishes this in closed form. The amplitude vanishes at zero discounting and is strictly increasing and strictly concave, so impatience steepens the crossing profile throughout, by less at each further increment.

Before stating the theorem we give the hypotheses something to hold of. The next subsection introduces four illustrative shapes a threshold can take, reads p, k and ℓ off each of them, and shows that the first two vary independently.

2.3 Four canonical geometries: fold, Hopf, pitchfork, transcritical

Figure 1 shows what the four illustrative geometries have in common. While the budget lasts the state rests in a valley and the boundary sits on a ridge some distance away, so crossing the threshold requires the noise to push the state over the ridge. Spending the budget moves the valley and the ridge toward each other. Every geometry closes its margin this way, so what separates them is the rate.

Left panels show the landscape $U(x;\mu)$ of each geometry against the state, with the stable equilibrium marked by a circle and the crossing boundary by a square, drawn at three budget levels decreasing from light to dark. As the budget falls the valley and the ridge move toward each other and the margin between them contracts. Panels A (fold) and B (pitchfork) have a margin that collapses as $\sqrt{\mu}$, panels C (transcritical) and D (Hopf) one that collapses linearly. D is a phase plane in which the shrinking limit cycle approaches the crossing circle.

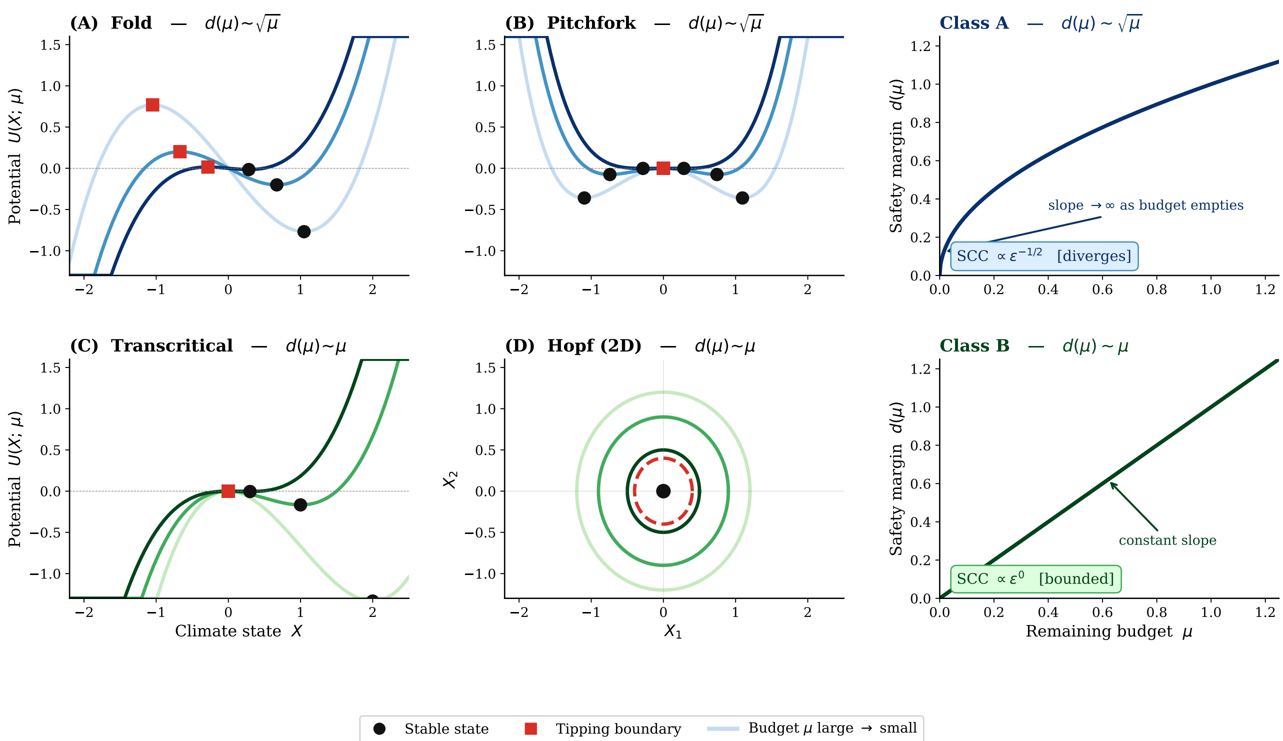

Figure 1: The four illustrative geometries.

The right panels show the resulting margin laws. The margin itself reaches zero at the threshold in every case, and what differs is its slope, the margin destroyed by the marginal tonne. Where $d(\mu) \sim \sqrt{\mu}$ the margin reaches zero with a vertical tangent, so the last tonne destroys unboundedly more margin than the first; where the slope is constant, every tonne destroys the same margin. That distinction, and not the type of bifurcation, is what decides whether the price diverges.

Table 1 reads the closing rate p, the local order k and the noise-layer width off each of the four normal forms. The first two rows are one dynamical system written twice: the fold $f = \mu - x^2$ and the transcritical $f = \mu x - x^2$ differ by the translation $x \mapsto x - \mu/2$, so both leave $-x^2$ at the threshold and share $k = 2$, and with it the noise layer and the amplitude. What separates them is the rate at which the budget closes the margin, $d = 2\sqrt{\mu}$ against $d = \mu$. The pitchfork $f = \mu x - x^3$ does the reverse, leaving $-x^3$ and therefore $k = 3$ while sharing the fold's closing rate.

Table 1: The four illustrative geometries

Geometry	Drift $f(x;\mu)$	Stable, boundary	$d(\mu)$	p	k	$\ell(\sigma)$
Fold	$\mu - x^2$	$+\sqrt{\mu}$, $-\sqrt{\mu}$	$2\sqrt{\mu}$	$\frac{1}{2}$	2	$\sigma^{2/3}$
Transcritical	$\mu x - x^2$	μ, 0	μ	1	2	$\sigma^{2/3}$
Pitchfork	$\mu x - x^3$	$\pm\sqrt{\mu}$, 0	$\sqrt{\mu}$	$\frac{1}{2}$	3	$\sigma^{1/2}$
Hopf	2-d	$r^*(\mu)$, r_c	$(\epsilon - \epsilon_c)/(2r_c)$	1	—	σ

That independence is what makes the theorem substantive. The closing rate will turn out to decide whether the price blows up as the threshold nears, while the local order decides how the noise enters, so a classification by one cuts across a classification by the other. Two thresholds can therefore be indistinguishable in their dynamics and still be priced on opposite sides of the split. It also means the classification cuts across the way bifurcations are usually grouped, since the fold and the transcritical share every local feature and fall on opposite sides, while the fold and the pitchfork differ in local order and fall together. For the pitchfork and the Hopf, (H2) holds only in the weaker form just noted, so the two conclusions of the theorem hold for them but are checked directly on their explicit amplitudes in Appendix A.8.1 rather than obtained as corollaries. The Hopf is also the one row whose state has two components, an oscillation rather than a resting point, which is why the local order has no counterpart there.

Section 3 places the Atlantic overturning circulation, the Amazon and the two ice sheets in the first row as folds, so that is the geometry carrying the empirical content of this paper.

2.4 The theorem

The next result gives the marginal value of control in closed form near the threshold. It shows that the two key features of this value, whether it diverges and how it responds to noise, are fixed by the closing rate of the margin alone.

Theorem 1. *Assume* (H1)–(H3)*, and let* Π *be the fixed-budget price* (3)*, built from the fixed-budget crossing value* φ*. Then, across the fixed-budget regime, by which we mean the range* *of proximities on which the budget barely moves while the state crosses and the stable state still sits inside the noise layer, and as ϵ decreases through it,*

$$\Pi(\epsilon,\sigma,\rho) = \mathcal{L}\,\underbrace{c\,p\,(M^*)^{p-1}}_{\text{geometry}}\;\underbrace{\frac{\Phi(\tilde\rho)}{\ell(\sigma)}}_{\text{steepness }\kappa(\sigma,\rho)}\;\underbrace{\epsilon^{p-1}}_{\text{proximity}}\,\big(1+o(1)\big), \qquad \underbrace{\tilde\rho = \frac{\rho}{m(\sigma)}}_{\text{rescaled discount rate}}\ . \tag{6}$$

Moreover:

(i) Dichotomy: across this regime the price follows the power law ϵ^{p-1}, so it grows without bound as the budget runs down when $p<1$ and stays bounded when $p \geq 1$.

(ii) Noise reversal: $\partial\Pi/\partial\sigma < 0$ throughout the fixed-budget regime, at every $\rho > 0$.

One qualification belongs with the statement. The regime it holds across is bounded at both ends: below by the dynamic scale ϵ_{dyn} of Section 4, beneath which the budget no longer stands still while the state crosses, and above by the proximity at which the stable state leaves the noise layer. The price is computed while the budget is held fixed, which is why we call it the fixed-budget price, and that approximation degrades in the immediate neighbourhood of the threshold, where the boundary itself moves appreciably as the state approaches it and the true price levels off rather than diverging[2].

For the upper end, lowering the noise at fixed proximity pushes the stable state out of the layer, since the layer narrows faster than the margin does: the very operation the reversal describes eventually carries the system beyond the zone where a marginal push can still change what happens. Beyond that point the crossing value is no longer close to one but exponentially small, and differentiating it in the budget gives a price that falls to zero as the noise vanishes rather than rising: the sign turns. Where it turns is worth locating, and the margin measured in layer widths is the scale to use, one width marking the layer's own edge. Numerically the sign turns at 0.94 widths for the overturning circulation, that is, just inside the layer rather than at its edge. Sitting inside the layer is therefore necessary for the reversal but not sufficient. The reversal is a statement about a window, not about every noise level.

The price is the budget derivative of the crossing value, as defined in equation (3), so the proof evaluates that derivative in three steps.

The first step asks how fast the crossing value falls as the state retreats from the boundary. Start at the boundary, where crossing is immediate and the claim is worth one, and move inward to the stable state a distance $d(\mu) = x^*(\mu) - x_-(\mu)$ away. Since φ solves a linear second-order equation with smooth coefficients it is smooth up to the boundary, so it can be expanded there,

$$\varphi\big(x^*(\mu);\mu\big) = \underbrace{\varphi(x_-)}_{=1} + \varphi'(x_-)\,d(\mu) + O\big(d(\mu)^2\big) = 1 - \kappa(\sigma,\rho)\,d(\mu) + o\big(d(\mu)\big), \qquad \kappa \equiv -\varphi'(x_-).$$

The crossing value falls as the state retreats, so $\varphi'(x_-) < 0$ and $\kappa > 0$: the shortfall below certainty is positive and proportional to the margin at leading order, and it is that shortfall the price differentiates.

[2]Pricing this calls for a dynamic treatment of the moving boundary, which we leave to future work.

Keeping only the linear term in the expansion above is legitimate only if the step taken is small relative to the scale on which φ changes, and that scale is the width of the noise layer $\ell(\sigma)$. What has to be small is therefore not the margin itself, since a margin of one degree means nothing until one knows how far the state typically wanders, but the margin measured against that width. More precisely half the margin, since the expansion is taken about the merged point and the two equilibria sit symmetrically at distance $d/2$ on either side of it. The quantity that has to be small is therefore

$$\eta^* = \frac{d(\mu)}{2\,\ell(\sigma)}, \tag{7}$$

which for the fold is $\sqrt{\mu}/\ell(\sigma)$, the equilibria sitting at $\pm\sqrt{\mu}$. It is the only dimensionless measure of distance the near-threshold problem admits. As the budget empties the margin closes while the layer width stays fixed, so η^* falls toward zero and one term suffices.

The first step gave the form of the shortfall but not the value of κ, and the rescaling supplies it. A slope is a change per unit of state, so its size depends on how the state is measured, and the near-threshold problem offers one length: the width of the noise layer. Write $x = \ell(\sigma)\,\eta$, so that η counts layer widths. Section 2.2 makes that substitution, and the noise level drops out, leaving an equation in $\tilde{\rho}$ alone. That equation does not determine its own solution, decay away from the threshold selecting a family whose members are multiples of one another, so the only feature of it the price can use is one free of that constant: the logarithmic slope, evaluated at the merged point about which the two equilibria sit symmetrically. We call it the amplitude, $\Phi(\tilde{\rho})$, and Appendix A.3 shows it to be positive. Reading it at the merged point rather than at the boundary is what makes it a function of $\tilde{\rho}$ alone: at a boundary sitting η^* widths away the slope depends on η^* as well, and the two agree to first order because η^* falls toward zero as the budget empties. Since $x = \ell\eta$, a slope in x is a slope in η divided by ℓ, so

$$\kappa(\sigma, \rho) = \frac{\Phi(\tilde{\rho})}{\ell(\sigma)} > 0.$$

Everything the noise and the discount rate do to the price passes through this one factor.

The second step asks how that shortfall moves as the budget shrinks, since the price is its rate of change. The shortfall is $\kappa\, d(\mu)$, and κ can be treated as constant: it is the slope of a function that does depend on the budget. Differentiating therefore differentiates the margin alone, and by (H1) that margin closes as $c\,\mu^p$, so

$$\left|\tfrac{\partial}{\partial\mu}\,\varphi\big(x^*(\mu);\mu\big)\right| = \kappa\, c\, p\, \mu^{p-1}(1 + o(1)).$$

Writing this in the remaining-budget fraction through $\Pi = \mathcal{L}\,|\partial_\mu\varphi|$ and $\mu = M^*\epsilon$ gives the master formula, in which the budget appears only through ϵ^{p-1}.

The third step turns to the noise. In the master formula only κ contains σ, since neither the geometry factor nor ϵ^{p-1} does, so one derivative settles the question. Write $\ell(\sigma) = \sigma^a$ and $m(\sigma) = \sigma^b$ for the two scales of (H2), with $a, b > 0$, so that $\kappa = \Phi(\rho/m)/\ell$. Raising the noise acts through both scales, and the two effects point the same way: the explicit factor $1/\ell$ shrinks, and the argument $\tilde{\rho} = \rho/m$ falls, which lowers Φ because Φ is increasing. Differentiating $\kappa = \sigma^{-a}\Phi(\rho\sigma^{-b})$ in σ gives two terms. The explicit factor contributes $-a\,\sigma^{-a-1}\Phi(\tilde\rho)$, and the argument contributes $\sigma^{-a}\Phi'(\tilde\rho)\cdot(-b\rho\sigma^{-b-1})$ by the chain rule. Rewriting $\rho\sigma^{-b-1}$ as $\tilde\rho/\sigma$ in the second and factoring out $\sigma^{-a-1} = 1/(\sigma\ell)$,

$$\frac{\partial\kappa}{\partial\sigma} = -\frac{1}{\sigma\,\ell(\sigma)}\Big[\,a\,\Phi(\tilde\rho) + b\,\tilde\rho\,\Phi'(\tilde\rho)\Big] < 0. \tag{8}$$

the bracket being positive because a and b are positive by construction and (H3) makes Φ and Φ' positive. So the steepness falls as the noise rises, and so does the price, at every value of σ and ρ for which the pair lies in the regime of the statement.

The two results are independent of one another. The price is a product of factors that live in different variables. One of them, ϵ^{p-1}, says how fast the threshold is approaching, and it carries the divergence: a square-root margin has a slope that blows up as the budget empties, whereas a linear margin has a constant slope and leaves the price finite. The other, κ, says how steep the crossing profile is, and it carries the reversal, through the two channels of (8). Because one factor lives in the budget and the other in the noise, neither constrains the other, which is why the reversal holds in every class while the divergence does not.

Table 1 now verifies the theorem. (H1) is read off the margin column, and (H3) is established in Appendix A.3 at every local order at once, so both hold in every row. (H2) is the one hypothesis whose status differs: noise and impatience collapse into $\tilde\rho$ exactly for the two rows with $k = 2$, and only approximately for the pitchfork and the Hopf. The two rows with $p = \frac{1}{2}$ therefore diverge and the two with $p = 1$ stay bounded, and in all four the price falls as noise rises.

The reversal is a statement about the marginal value of restraint, and not about the total risk of crossing. Fix the proximity ϵ and raise the noise. The marginal price falls, which is what $\partial\Pi/\partial\sigma < 0$ says. At the same time the total expected loss from crossing rises, because that loss is $\mathcal{L}\varphi$ and more noise makes crossing both more likely and sooner. A noisier system is therefore more exposed overall while each extra unit of restraint buys less. Nothing forces the two to move together, since one of them is a level and the other is a derivative.

2.5 Scope of the results

Three limits are worth stating. The first two ask whether the divergence exponent could be an artefact of how the problem is written rather than a property of the system; neither is. The last bound the region over which the result holds.

2.5.1 The coordinate the state is written in

The crossing time is a property of the path and not of the coordinate used to describe it. Relabel the state by any strictly increasing $Y = \chi(X)$ whose slope is finite and positive at the merged point. The events $\{X_t \le x_-\}$ and $\{Y_t \le \chi(x_-)\}$ are then the same event, so τ is the same random time in both descriptions and

$$\varphi\big(x^*(\mu);\mu\big) = \mathbb{E}_{x^*}\big[e^{-\rho\tau}\big] = \mathbb{E}_{y^*}\big[e^{-\rho\tau}\big] = \varphi_Y\big(y^*(\mu);\mu\big).$$

The price is the budget derivative of this one scalar function, so the price, the divergence exponent and the sign of the noise response are the same in every coordinate. The margin rescales by χ' at the merged point, which moves the constant c and leaves the closing rate untouched, while the layer width and the amplitude each change and their product does not. The closing rate is therefore an invariant of the dynamics, and classifying an element does not depend on whether it happens to be measured in degrees or in a vegetation index. The same identity settles the structure of the noise. Additive noise in one coordinate is state-dependent noise $\sigma g(X)\,dW$ in another, the two describe the same paths, and they price the edge identically provided g neither vanishes nor blows up at the threshold.

2.5.2 Preferences

The two theorems are proved under log utility, and the question is whether they survive when risk aversion is freed from it (complete proof in A.6).

Epstein–Zin at unit elasticity keeps $c^* = \rho K$, a fixed fraction of wealth free of the climate state, so the reduction survives with one term added, a penalty on the variance of the climate state. That term lives on the same length scale as the diffusion, so it reshapes the amplitude and leaves untouched the rate at which the edge approaches, which is set by the dynamics and not by tastes: the exponent is derived exactly as under log. The reversal survives on a weaker footing, since it rests on two properties of the reshaped amplitude, of which we prove one and verify the other numerically across a range of impatience. Power utility ties risk aversion to the elasticity, and optimal consumption is then still proportional to wealth but with a factor that moves with the climate state, so the reduction is lost. Both theorems have to be computed rather than derived, and both hold.

Neither theorem needs the log: what log utility buys is not the results but the closed form.

2.5.3 Where the calibrated elements sit

None of the three elements is inside that window today, for two reasons. First, the overturning circulation and the Amazon are too far from their thresholds as analyzed from their own stability. An element sitting close to its edge, measured against the size of its own fluctuations, does not stay put: the noise carries it across within a few recovery times. Both have held their state for very much longer than that, so both must be far. Second, the West Antarctic ice sheet is too slow instead. The fixed-budget price holds while the budget drains slowly relative to the system's own recovery, which asks that an element recover faster than its deadline arrives. The ice sheet recovers over centuries against a deadline of decades, so the static law does not describe it at any proximity. Its slowness is what spares it from the first argument and what excludes it under the second.

Emissions close the first gap. Proximity falls as the budget is spent, so an element outside the window today enters it strictly before its own deadline arrives. What the paper prices is the regime these elements will occupy, and what it dates is the arrival of the deadline within that regime. The exponent, the sign of the noise effect and the classification are properties of the near-threshold problem and hold wherever that problem describes the system.

3 Application to climate

The general theorem was stated for any system in which a controllable drift pushes a state toward a collapsing boundary. Climate is such a system, and this section reads the theorem in climate terms. It does three things: it instantiates the general price as an explicit formula for a fold element, it identifies the observable that says whether a given element belongs to the divergent class or the bounded one, and it shows that the object being priced is the value of postponing a crossing rather than of avoiding it, which reverses the familiar comparative static in the discount rate.

3.1 The threshold component of the social cost of carbon

This subsection instantiates the general price for a fold. Each of the climate tipping elements we study is one: as warming rises, a stable climate state and an unstable one approach, merge at a threshold, and vanish. We take that geometry as given and price it. The first row of Table 1 supplies everything the theorem needs, with closing rate $p = \frac{1}{2}$, constant $c = 2$, and noise-layer width $\ell = \sigma^{2/3}$, so the hypotheses hold and the two conclusions read, in climate terms, that the threshold component of the SCC diverges as the budget to the threshold is exhausted and that it rises as the forcing on that element grows quieter.

Two changes adapt the general price to a climate element. First, the budget is the carbon budget to the threshold, so the state carrying the system toward its edge is the remaining-budget fraction $\epsilon = \mu/M^*$ and the constant M^* is that budget. Second, the loss is not incurred all at once: an element realizes its damage over a post-tipping timescale T, decades for the AMOC and millennia for an ice sheet. Spreading $\mathcal{L}$ over that horizon and discounting it back to the crossing instant replaces $\mathcal{L}$ by $\mathcal{L}\,\mathsf{f}(\rho T)$, where $\mathsf{f}(\rho T) = (1 - e^{-\rho T})/(\rho T)$ is the average discount factor of a flow of duration T, equal to one for an instantaneous loss and falling toward zero as the damage is pushed further out.

3.1.1 The fold price in closed form

Substituting the fold row into the master formula (6) then gives the price in closed form. The geometry factor is $c\,p\,(M^*)^{p-1} = 1/\sqrt{M^*}$, the steepness is $\Phi_{\text{fold}}(\tilde{\rho})/\sigma^{2/3}$ with $\tilde{\rho} = \rho/\sigma^{2/3}$, and the proximity factor is $\epsilon^{-1/2}$, so

$$\mathrm{SCC}_{\text{threshold}}(\epsilon) = \frac{A}{\sqrt{\epsilon}}, \qquad A \equiv \mathcal{L} \cdot \mathsf{f}(\rho T) \cdot \frac{\Phi_{\text{fold}}(\tilde{\rho})}{\sigma^{2/3}\sqrt{M^*}}. \tag{9}$$

The formula reads in two parts. The constant A is the level of the premium, set by how bad crossing is through $\mathcal{L}$, how slowly the damage arrives through $\mathsf{f}(\rho T)$, how large the budget is through M^*, and how noisy the element is and how impatient society is through $\Phi_{\text{fold}}(\tilde{\rho})/\sigma^{2/3}$. The factor $\epsilon^{-1/2}$ is the shape of the divergence, identical for every fold element. How the damage is realized after crossing enters only the level through $\mathsf{f}(\rho T)$ and never the budget dependence, since the singularity is fixed by how fast the margin collapses and not by how the damage unfolds.

That split governs what the formula can and cannot deliver here. Anything that is a ratio in ϵ comes from the second part alone and is available at once, which is what the rest of this section uses. A level in dollars per tonne comes from the first, and it is not available yet: the noise level is estimated in the element's own observed variable, in degrees or in an index, while the budget is in gigatonnes, and the normal form uses neither. Closing that gap takes one measurement per element, the distance from the current stable state to the unstable one expressed in the variable the element is actually observed in. Section 4 shows why no element has it and reports the level as a family indexed by it.

3.1.2 When the premium can be added to a conventional price

The threshold component adds to the ordinary Ramsey component without a cross-term, so a practitioner can bolt (9) onto an existing integrated model rather than replacing it,

$$\mathrm{SCC}_{\mathrm{total}} \;=\; \mathrm{SCC}_{\mathrm{Ramsey}} \;+\; \mathrm{SCC}_{\mathrm{threshold}}(\epsilon)\Big[1 + O\Big(\tfrac{E_0}{\rho M^* \epsilon}\Big)\Big]. \tag{10}$$

The practitioner reads $\mathrm{SCC}_{\mathrm{Ramsey}}$ off the existing model, computes $\mathrm{SCC}_{\mathrm{threshold}}$ element by element and adds the two. Where elements interact the sum becomes a lower bound, and Appendix A.9 treats the two-element case.

The split is not automatic, since emissions move the consumption path and the distance to threshold at once. Under logarithmic utility the two channels come apart cleanly, and the only thing left over is that the crossing value itself drifts as the budget drains: a price computed at a frozen budget misses that drift. Appendix A.4 bounds what it misses, and the error is the fraction of the budget a planner burns through in one discounting time, divided by the fraction still unspent. The split is therefore usable while the remaining budget is large relative to what impatience consumes of it, and it is exact once emissions stop.

It binds unevenly across the three elements. Evaluated at current proximities, the relative error is 21% for the overturning circulation and 26% for the Amazon, so the split is usable there with an error a practitioner can carry. For the WAIS it exceeds one, the correction being larger than the term it corrects, so the split does not hold at all. The failure rules out for the ice sheet is bolting a threshold term onto an existing model, which the other two admit.

3.1.3 A price path steeper than Hotelling

The price also implies a path, and its shape differs from the one a planner expects. Along a trajectory where the budget drains at a constant rate, $\dot{\epsilon} = -E_0/M^*$, we have $\log \mathrm{SCC}_{\mathrm{threshold}} = \log A - \frac{1}{2}\log \epsilon(t)$, so

$$\frac{d}{dt}\log \mathrm{SCC}_{\mathrm{threshold}} = \frac{E_0}{2M^*\epsilon(t)}, \qquad \frac{d^2}{dt^2}\log \mathrm{SCC}_{\mathrm{threshold}} = \frac{E_0^2}{2(M^*)^2\epsilon(t)^2} > 0.$$

A standard externality follows a Hotelling path: the tax grows at a constant rate. Here the growth rate rises as the threshold nears: the optimal tax grows faster than exponentially.

3.2 Two classes, and the observable that separates them

Which climate elements have a divergent price and which have a bounded one is settled by the rate at which the margin closes, as Theorem 1 shows. This subsection asks whether that assignment can be tested against data rather than inferred from structural models, and identifies a statistic that can do so.

Two candidates fail first. The noise reversal cannot separate the classes, since it holds in every geometry. Nor can how strongly the price responds to the noise level. Noise enters the price twice in (6), and only through the steepness κ: through the width of the layer $\ell(\sigma) = \sigma^a$ in the denominator, and through the rescaled discount rate $\tilde{\rho} = \rho/m(\sigma)$ in the argument of the amplitude, with $m(\sigma) = \sigma^b$. Differentiating $\log \Pi$ in $\log \sigma$ therefore gives $-(a + bE)$, where E is the elasticity of the amplitude in $\tilde{\rho}$: the first channel always counts in full, the second only in proportion to how much the amplitude responds.

Carrying that through, the elasticity is $-2k/(k + 1)$ where impatience is small relative to the layer, which is where the calibrated elements sit: it reads the local order of the drift at the threshold and nothing else. That is enough to disqualify it. The fold and the transcritical share the same local order, so both give $-4/3$ while sitting in opposite classes, and the pitchfork gives $-3/2$, so the statistic varies inside a single class as well. The two features are independent, as Table 1 showed: the elasticity reads the shape of the drift, the class the rate at which the margin closes.

What does separate the classes is not the sensitivity itself but the way it changes as an element approaches its threshold. Formally this is the cross-partial of the log price in noise and proximity, $\partial^2 \log \mathrm{SCC}_{\text{threshold}}/\partial \log \sigma\, \partial \log \epsilon$. Take two elements at the same distance to threshold, one driven by weak noise and one by strong noise, and form the ratio R of their prices. The reversal says that ratio exceeds one. The cross-partial says whether it widens or narrows as the elements are pushed closer to their thresholds. We state the result on R, which is what the numerical solution plots below. The two are related by

$$\frac{\partial \log R}{\partial \log \epsilon} = -\big(\log \sigma_{\text{high}} - \log \sigma_{\text{low}}\big)\, \frac{\partial^2 \log \mathrm{SCC}_{\text{threshold}}}{\partial \log \sigma\, \partial \log \epsilon}, \tag{11}$$

so a ratio that widens with the remaining budget is a negative mixed derivative. That the two classes differ at all in this quantity follows from the theorem. The leading price is a product of a factor in the noise and a factor in the budget, so in logarithms the two add and the mixed derivative kills it: the effect lives in the correction to that leading price. That correction is of relative order η^*, the margin measured against the width of the noise layer, which is proportional to ϵ^p. The closing rate that fixes the divergence exponent therefore fixes the signature as well, and no further element-specific input enters.

The diagnostic separates the classes, and it separates them better the nearer the threshold. On the pair of noise levels of Figure 2, the cross-partial in the divergent class exceeds the bounded-class one by a factor of about 6.0 at $\epsilon = 0.15$ and about 125 at $\epsilon = 0.05$: the gap widens as an element approaches its edge. The figure shows the predicted split. The fold and the pitchfork depart upward from the common asymptote as the remaining budget grows, the fold fastest, while the transcritical stays within 8% of it[3]. The comparison is numerical: for each geometry the discounted crossing value $\varphi(x;\mu) = \mathbb{E}_x[e^{-\rho\tau}]$ solves $\frac{\sigma^2}{2}\varphi'' + f(x;\mu)\varphi' = \rho\varphi$ on $(x_-(\mu), \infty)$ with $\varphi(x_-) = 1$ and $\varphi(\infty) = 0$, and the price is read off $\mathrm{SCC}_{\text{threshold}} \propto |\partial\varphi/\partial\epsilon|$, with numerical details in Appendix A.7.

[3]The Hopf is left off the grid. Its threshold is reached at a strictly positive proximity rather than at zero, and its radial equation carries an extra term, so it does not sit on the same axes (Appendix A.8.1).

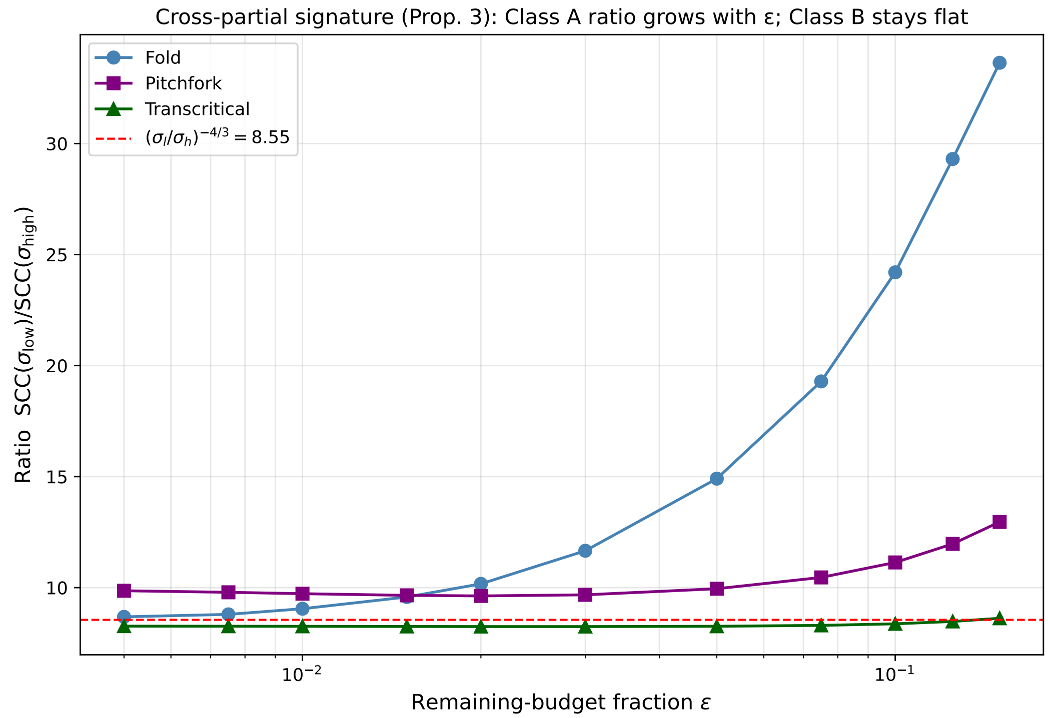

Figure 2: Cross-partial signature. The ratio $\mathrm{SCC}_{\mathrm{threshold}}(\sigma = 0.3)/\mathrm{SCC}_{\mathrm{threshold}}(\sigma = 1.5)$ is plotted against the remaining-budget fraction ϵ (log scale on the horizontal axis).

Table 2 collects geometries with their class and the climate elements that realize it.

Table 2: Four canonical geometries with their classes and climate elements.

Bifurcation	Class	Climate examples	Pricing implication
Fold	A	AMOC, Greenland, West Antarctic, Amazon	Diverges as $\epsilon^{-1/2}$; finite deadline
Pitchfork	A	None[a]	Same exponent as the fold
Transcritical	B	Smooth regime shifts	Bounded; standard externality
Hopf	B	ENSO-type oscillations	Bounded

The list is illustrative. AMOC: bistability robust from thermohaline box models (Cessi, 1994). Greenland: multistability and hysteresis in ice-sheet models (Robinson et al., 2012). West Antarctic: marine ice-sheet instability yields a fold (Schoof, 2007). Amazon: moisture-recycling bistability supported (Staal et al., 2020), though whether a basin-wide threshold exists is debated. ENSO: the conceptual model of Jin (1997) is a two-variable oscillator that switches from noise-driven to self-sustained as the coupling strengthens, which is a Hopf; whether ENSO is a tipping element at all is contested. Other candidates are left unclassified: Armstrong McKay et al. (2022) treat boreal permafrost as a strong but non-runaway feedback with limited large-scale bistability, so it would sit in Class B if anywhere.

[a] The pitchfork enters here as an internal check that the divergence exponent is set by the closing rate alone, independently of the $\sigma^{1/2}$ noise scale that distinguishes it from the fold.

Two limits on what this statistic can do are worth stating.

The first is that it orders proximities within a single element and not elements against one another. Within the divergent class the reversal operates element by element: for a given element, weaker noise raise its own premium at a fixed distance to threshold. Ranking two elements by their noise levels is another matter, and the obstacle is one of units. Each noise level is estimated on that element's own observed variable, and those variables live in different physical spaces: kelvin of sea-surface temperature for the AMOC, a dimensionless vegetation index for the Amazon, ocean-forcing units for the West Antarctic ice sheet. Making them comparable would take a common currency, such as the movement each element's noise induces in the distance to its own threshold, and that needs a mapping from the observed variable to that distance which exists for none of the three.

What differs across the three is how much else is in hand. The AMOC has both a measured recovery rate and a measured variance, which is why its price can be written as an explicit family indexed by the margin. The Amazon has a recovery rate, but in an index whose scale is not tied to its threshold. The West Antarctic has neither, its relevant variability lying beyond the instrumental record. What is common is what the geometry fixes without reference to any conversion: the exponent, the sign of the noise effect, the cross-partial itself, and the date at which abatement should switch on.

The second limit is that the instrumental record cannot run this test, and the obstacle is not the length of the record. The cross-partial requires the noise level to vary at controlled proximity, and neither variation is available: the noise level of a given element does not move over the record, and noise levels are not comparable across elements.

A designed ensemble of simulations meets all conditions, which is where we propose the test be run. The noise level is a parameter one sets rather than a quantity one estimates, the proximity is known by construction and can be placed where the contrast is strong, the prices are internal to the model and need no conversion, and holding the budget fixed is the approximation the price is computed under. The design is a single element in a single model, two or three levels of stochastic forcing, and a sweep of proximities. The prediction is that the price ratio widens with the remaining budget throughout, and that the contrast between a divergent and a bounded geometry grows as the threshold is approached. What the experiment does require is that the distance to threshold be adjustable, so the experiment belongs to models of intermediate complexity in which that distance is a parameter, and not to general circulation models in which it is neither known nor tunable.

3.3 A price for delay, not for avoidance

The literature prices crossing a threshold through a hazard rate, so translating our price into that language is a way to highlight the difference. Suppose crossing arrived at a constant rate h per unit of time, as in a Poisson model. The date of crossing would then be exponentially distributed with mean $1/h$, and the present value of one unit paid at that date:

$$\varphi = \mathbb{E}\big[e^{-\rho\tau}\big] = \frac{h}{\rho + h}.$$

The relation can be read in the other direction. Given any crossing value, we can ask what constant rate would have been needed to produce it, and inverting gives that number, $h = \rho\,\varphi/(1-\varphi)$. This is an implied hazard in the same sense as an implied volatility: not a rate that anything in our model follows, but the rate a Poisson model would need in order to agree with the price we compute.

Applying it to ours is immediate, and the result is the price read backwards. Near the threshold the crossing value is close to one, so the denominator carries everything and $h \approx \rho/(1-\varphi)$. The shortfall $1-\varphi$ of Appendix A.5.1 is, up to the loss and the horizon factor, exactly what (9) prices, so

$$h(\epsilon) \;=\; \frac{\rho\,\mathcal{L}\,\mathsf{f}(\rho T)}{2\,M^{*}} \cdot \frac{1}{\mathrm{SCC}_{\mathrm{threshold}}(\epsilon)}.$$

The implied hazard is inversely proportional to the price, with a constant free of the noise level, of the amplitude and of the proximity. It therefore grows without bound as the budget empties, at exactly the rate the price does.

It matters what that explosion is and is not. It is not a rising probability that the system tips, because at any fixed budget the boundary sits a finite distance away and the noise eventually carries the state across it, so crossing occurs in finite time almost surely and its probability is already one. Since $1/h$ is the mean waiting time, what diverges is the reciprocal of a delay: the implied hazard rises because crossing becomes imminent, not because it becomes likely. The crossing value falls short of one only through discounting, which is why the shortfall $1-\varphi$ is the quantity the price differentiates.

Thus, restraint never buys avoidance, since the system tips whatever the planner does at the margin, and it buys only postponement. The amplitude records this exactly. Consider a planner who does not discount, so that $\tilde{\rho} = 0$. Such a planner is indifferent between a certain loss now and the same certain loss later, so postponement is worth nothing, one more unit of restraint is worth nothing, and the price is zero. This is $\Phi_{\mathrm{fold}}(0) = 0$.

Postponement is worth something only to a planner who discounts, and it is worth more to a planner who discounts more, because moving a loss one year further out gains more the heavier the discounting. The threshold component of the price therefore rises with ρ, where the ordinary Ramsey component falls: the Ramsey price discounts a future damage, which a higher ρ makes lighter, whereas ours prices a delay, which a higher ρ makes dearer. The exponent is unaffected, but the level is not, and the amplitude says by how much. Across the Nordhaus and Stern range $\rho \in [1\%, 5\%]$ the amplitude Φ_{fold} varies by a factor of 4.7, since it is close to linear in $\tilde{\rho}$ at the small values the climate case occupies (Appendix A.7.2). Impatience reaches the price a second time, however, and in the opposite direction. The loss arrives over the realization time T rather than at once, so the price carries $\mathsf{f}(\rho T)$ alongside $\Phi_{\mathrm{fold}}(\tilde{\rho})$, and a higher ρ discounts a damage that unfolds slowly all the more heavily.

The two channels therefore net out differently from one element to the next, and what decides which of them wins is the realization timescale. Over the same discount range the tipping price rises by a factor of 2.17 for the overturning circulation, whose damage is realized over 50 years, by 1.46 for the Amazon at 100 years, and is flat for the WAIS at 2000 years, where the ratio is 0.93. The unresolved disagreement about discounting therefore moves the two components of the carbon price in opposite directions, by amounts that depend on the element, which is not how that debate is usually understood.

The same reading sharpens the disagreement with the literature, and locates it precisely. Read through the implied hazard, each of the three representations of the introduction is an assumption about the delay, and each of them guarantees that some delay always remains. A hazard that stays finite at every finite stock keeps the expected wait bounded away from zero, so spending the budget makes crossing likely but never imminent. An unknown threshold location makes the wait something the planner infers rather than something the dynamics deliver, and it stays positive as long as the threshold might still lie further off. An expansion in small tipping risk needs the wait to stay long enough for that risk to remain a perturbation, so a vanishing delay would destroy the method rather than be a result of it. A collapsing margin leaves no such guarantee. The implied hazard is singular at a finite and identifiable stock, so the delay can be driven to zero by spending the budget. The disagreement is therefore about whether the hazard is smooth at the threshold, not about the pricing method, and substituting our $h(\epsilon)$ into a framework of the kind Cai and Lontzek (2019) use returns the singular price.

Section 4 calibrates the price for each element, turns it into dates, and reports how far those dates and prices are resolved by the data.

4 From price to policy

The formula (9) turns the theory into numbers for three fold-class elements, namely AMOC, the Amazon dieback, and the West Antarctic Ice Sheet. We treat AMOC in the most detail, because it is the one element whose σ can be anchored on an instrumental record. Greenland is also a fold, but we leave it out of the calibrated trio. Its realization timescale runs to several millennia, which makes it, like WAIS, dominated by the discounting factor $\mathsf{f}(\rho T)$, so the trio of AMOC, Amazon, and WAIS already spans the relevant damage and timescale regimes of circulation, biosphere, and ice without it.

The section does three things in turn: it calibrates the five observables for the climate elements and reports the resulting carbon price; it identifies which observable that price is most sensitive to; and it converts the price, through a backstop technology, into an endogenous decarbonization deadline.

4.1 Calibrating the formula

Calibrating (9) means fixing five observables for each element, and they do not all stand on the same footing. We take them in turn. The noise comes first, because it is the only one the climate datasets do not report and we have to construct. The damage comes second, because it is the one input we deliberately leave as a range rather than as a value. The three remaining parameters follow from the literature through a single conversion, and the subsection closes on the price they imply and the one number that price still needs.

4.1.1 The noise

Of the observables, the noise level σ is the only one no dataset reports, so it has to be recovered from an early-warning signature, and the three elements stand on very different footing when it comes to doing so. This subsection gives the estimator and takes the three elements in decreasing order of how well the estimator can be applied.

The estimator comes from critical slowing down. As a system approaches a fold it recovers ever more slowly from disturbances (Scheffer et al., 2009), and that slow recovery inflates the variance of its fluctuations in a way the geometry fixes completely. The recovery rate is the slope of the drift where the element rests. Push the element a small distance from $x^* = \sqrt{\mu}$: the drift is zero there, so it is well approximated by its slope, $f'(x^*) = -2\sqrt{\mu}$, and the displacement decays at rate $\lambda = 2\sqrt{\mu}$. As the budget empties the two equilibria move toward each other, the drift flattens between them, and the pull back toward the resting point vanishes at the threshold. The element is then an autoregression in continuous time, $dx = -\lambda x\, dt + \sigma\, dW$. Noise adds σ^2 to the variance each period, mean reversion removes the fraction 2λ of what has accumulated, and the two cancel at $\sigma^2/(2\lambda)$. A slower pull therefore means a larger variance at unchanged noise.

Reading that balance in the other direction recovers the noise from two quantities a record supplies,

$$\sigma = \sqrt{2\,\lambda\,\mathrm{Var}(X_t)}, \tag{12}$$

the recovery rate from the autocorrelation of the detrended series and the variance from the series itself. One assumption enters with it: reading λ off a record and setting it equal to $2\sqrt{\mu}$ takes the element observed to be the system the model describes, and it is what fixes the budget in the units the model is written in. What it buys is a rate and not a distance, which is why the exponent and the sign of the noise effect need it and nothing more, whereas a price in dollars needs the missing margin as well.

Early-warning indicators report rising variance as a threshold nears, which seems to contradict a reversal that makes quieter systems dearer. It does not, and the decomposition $\mathrm{Var} = \sigma^2/(2\lambda)$ says why: at fixed σ, the observed variance rises without bound because the restoring pull λ collapses, not because the forcing grows. What the indicators track is lost resilience; what the reversal is about is the forcing. The two are separate quantities and they push the price the same way, resilience through $\epsilon^{-1/2}$ and forcing through $\sigma^{-2/3}$.

For AMOC, equation (12) can be applied. On the Caesar et al. (2018) HadISST series (1871–2016) it gives $\hat{\sigma} \approx 0.260\,\mathrm{K\,yr^{-1/2}}$, which anchors the AMOC prior in Table 3. This is a noise level in the units of a physical observable, which is what makes the AMOC the one element whose price can be written as an explicit family indexed by the margin.

For the Amazon, the estimator applies but its output is not comparable. The early-warning literature documents resilience loss through the vegetation optical depth (Boulton et al., 2022), whose recovery rate is fast, roughly $9.7\,\mathrm{yr^{-1}}$. But the optical depth is a dimensionless index, so equation (12) returns a noise level in units of that index, with no defensible conversion to a temperature or a budget. Any estimate is therefore an effective noise of the observed vegetation dynamics, informative about the Amazon in its own units and not commensurable with the AMOC's. We treat the Amazon σ as a wide effective prior whose order of magnitude is constrained by the observed recovery rate rather than a single value.

For the West Antarctic ice sheet, even an effective estimate is out of reach, since the millennial realization time places the relevant variability far beyond the instrumental ice record. The physical driver is not the ice mass but the ocean forcing, and the decadal variability of

Amundsen Sea warming and basal melt is real and documented (Haigh and Holland, 2024; Jenkins et al., 2018), but translating that forcing into a noise level on the ice-sheet state requires a model-dependent sensitivity we do not attempt to calibrate. We therefore treat the West Antarctic σ as an explicitly epistemic prior, wide and not claimed to be observed.

4.1.2 The damage

$\mathcal{L} = (d/\rho)\,$GDP capitalizes a permanent annual flow loss; at $\rho = 3\%$ the range $[0, 0.67]$ corresponds to flows of roughly 0–2% of GDP per year. We draw $\mathcal{L}/$GDP uniformly over this range, common to all three elements, a robustness sweep rather than calibrated beliefs, thinner than previous stylized work (Fillon et al., 2023). The common prior is deliberate. Element-specific tipping damages are not yet settled enough to differentiate: where estimates exist at all, they typically rest on a single study per element, and the aggregate figures of Dietz et al. (2021) are built from that thin base rather than from a body of converging results. Thus, the cross-element differences in $\mathrm{SCC}_{\mathrm{threshold}}$ come from what the model identifies (geometry, proximity ϵ, and timescale T) not from damage assumptions. We floor $\mathcal{L}$ at zero. This matters mainly for the AMOC, whose only net-benefit sign traces to a single debated result in which a weaker circulation cools some regions (Anthoff et al., 2016); rather than import a negative damage resting on such ground, we let that element enter at zero. Damage sets the level of the price, never the exponent or the noise sign.

4.1.3 The remaining parameters, from the literature

Thresholds ΔT^* and realization times T are the central estimates and ranges of Armstrong McKay et al. (2022), and current warming is $\Delta T_0 = 1.2\,°\mathrm{C}$ (IPCC, 2021). Warming is converted into cumulative emissions through the IPCC AR6 transient climate response to cumulative emissions, drawn as $\mathrm{TCRE} \sim \mathcal{N}(0.45, 0.095^2)$ in degrees per $1{,}000\,\mathrm{GtCO_2}$ and truncated to $[0.14, 0.82]$, which is the interval more commonly quoted as $[0.5, 3.0]$ degrees per $1{,}000\,$GtC. The Monte Carlo propagates that uncertainty.

Two budgets follow from the conversion and they are not the same. The budget from pre-industrial to the threshold sets the scale of the problem,

$$M^* = \frac{\Delta T^*}{\mathrm{TCRE}} \times 1{,}000\ \mathrm{GtCO_2}, \tag{13}$$

whereas the part of it still unspent today is $\mu_{\mathrm{now}} = M^*\,\epsilon_{\mathrm{now}}$. The proximity is the ratio of the two, and since both carry the same conversion factor it reduces to a ratio of temperatures, $\epsilon_{\mathrm{now}} = \max(0,\ (\Delta T^* - \Delta T_0)/\Delta T^*)$, equal to one at pre-industrial, when the full margin is intact, and to zero at the threshold. For the AMOC, at a central threshold of $4\,°\mathrm{C}$, this gives $\epsilon_{\mathrm{now}} = (4 - 1.2)/4 = 0.70$.

Table 3 lists the five observables with their intervals and sources. We set $\rho = 0.03$ and draw $N = 100{,}000$ Monte Carlo samples over them, reporting medians and 90% credible intervals for the quantities the calibration determines.

Table 3: Calibration inputs for the three fold-class elements. Central value with [P10, P90] where a range is shown; ϵ_{now} is the current proximity.

Element	$\mathcal{L}$/GDP	ΔT^* (°C)	M^* (GtCO$_2$)	$\sigma^\dagger$	T (yr)	ϵ_{now}
AMOC	$[0, 0.67]$	$4^\ddagger$ $[1.4, 8.0]$	9158	0.260 $[0.18, 0.38]$	50 $[15, 300]$	0.70
Amazon	$[0, 0.67]$	3.5 $[2.0, 6.0]$	7908	0.15 $[0.06, 0.36]$	100 $[50, 200]$	0.66
WAIS	$[0, 0.67]$	1.5 $[1.0, 3.0]$	3828	0.15 $[0.06, 0.36]$	2000 $[500, 13{,}000]$	0.20

†Each σ is in the units of its own element's observable, so the column is not comparable across rows: the AMOC entry is estimated from the record in K yr$^{-1/2}$, the other two are dimensionless effective priors.
‡ The AMOC proximity is contested by an order of magnitude. We calibrate at the central threshold of Armstrong McKay et al. (2022). Reading the same record through early-warning indicators, Ditlevsen and Ditlevsen (2023) infer collapse within the century, which would place the element far nearer its edge.

4.1.4 The resulting price

The inputs above determine the price up to one number, which is the conversion between the units the model is written in and those the element is observed in. Two things about the price hold without it.

The first is a rate of increase. Since the price varies as $\epsilon^{-1/2}$, the ratio between any two proximities is fixed by the exponent alone: moving from $\epsilon = 0.30$ to $\epsilon = 0.10$ multiplies the tipping component of the carbon price by $\sqrt{3}$, for every element and whatever the level.

The second is a ranking of how heavily each element's damage is discounted. The realization factor $\mathsf{f}(\rho T)$ is a pure discount on the loss, so it is comparable across elements without any conversion: it equals 0.52 for the AMOC, 0.32 for the Amazon and 0.017 for the West Antarctic ice sheet. These two pull against each other for the ice sheet, which is much the closest to its threshold, at $\epsilon \approx 0.20$, and yet has its damage discounted by some two orders of magnitude more than the AMOC's. Which of them wins is what the deadlines of Section 4.2 settle.

What does not follow is a number in dollars per tonne, because the model and the record do not measure the state in the same units. The drift $f(x; \mu) = \mu - x^2$ already fixes its own: dimensional consistency leaves the state carrying the dimension of an inverse time, so the state of the model is neither a temperature nor a stock of carbon, and entering σ in kelvin and M^* in gigatonnes does not produce a price. One number per element is missing, the scale that carries a unit of the observed variable into the state of the model. We call it the physical margin d_{phys}, the distance from the current stable state to the unstable one in the variable the element is actually observed in: kelvin for the AMOC, a vegetation index for the Amazon.

That number is the only thing the record cannot give. It gives two others, and they are enough to reduce the price to a single unknown.

The first is the recovery rate. Push the element slightly away from where it rests and it comes back, the faster the steeper the drift there, at rate $\lambda = 2\sqrt{\mu}$ for the fold. The margin happens to be $2\sqrt{\mu}$ too, so in the model's units the recovery rate and the margin are the same number. That is what makes the record enough: a recovery rate does not depend on the units the state is measured in, so reading it off fixes the budget with no conversion.

The second is the year-to-year variance. Noise pushes the element away from where it rests and the drift pulls it back, and its variance settles where the two balance: noise puts variance in at rate σ^2 and the pull takes it out at rate 2λ, so $\sigma^2 = 2\lambda\,\mathrm{Var}$.

They combine because the price uses a ratio and not a distance: how large the margin is relative to how much the element fluctuates from year to year, which is the η^* of (7). Both are measured in the same variable, so the unknown conversion cancels between them, and the point of what follows is to rewrite that ratio in observed quantities rather than in the model's. The layer width is $\ell = \sigma^{2/3} = (2\lambda\,\mathrm{Var})^{1/3}$ by (12), and the margin is $d = 2\sqrt{\mu} = \lambda$ by the identification of Section 4.1.1, so

$$\eta^* = \frac{d}{2\ell} = \frac{\lambda}{2\,(2\lambda\,\mathrm{Var})^{1/3}},$$

The variance in that expression is still in the model's units, and the record is not. Let c convert between them, one unit of the model's state being c units of whatever the element is observed in. A distance scales by c and a variance, being a squared distance, by c^2: the observed margin is $d_{\mathrm{phys}} = c\lambda$ and the observed variance is c^2 times the model's. The combination $d_{\mathrm{phys}}^2/\mathrm{Var}$ therefore carries c^2 in both places and does not depend on c at all, which leaves

$$\eta^* = \tfrac{1}{2}\left(\frac{d_{\mathrm{phys}}^2}{2\,\mathrm{Var}}\right)^{1/3}, \tag{14}$$

entirely in quantities the element is observed in.

With η^* in hand the price depends on the element through the recovery rate and the margin alone,

$$\mathrm{SCC}_{\mathrm{threshold}} = 4\,\mathcal{L}\,\mathsf{f}(\rho T)\,\Phi_{\mathrm{fold}}(\tilde{\rho})\,\frac{\eta^*}{\lambda^2}, \qquad \tilde{\rho} = \frac{2\rho\,\eta^*}{\lambda}, \tag{15}$$

which is (9) rewritten: substituting $\sqrt{M^*\epsilon} = \sqrt{\mu} = \lambda/2$ and $\ell = \sqrt{\mu}/\eta^*$ into it gives the displayed form, and the same two substitutions turn $\tilde{\rho} = \rho/\ell$ into the expression alongside. For the AMOC, with $\lambda \approx 0.49\,\mathrm{yr}^{-1}$ and $\mathrm{Var} \approx 0.068\,\mathrm{K}^2$ from the HadISST series, and at the median of the damage prior, this gives a one-parameter family,

$$\mathrm{SCC}_{\mathrm{threshold}} \approx \$270 \times \left(\frac{d_{\mathrm{phys}}}{\mathrm{SD}}\right)^{4/3} \text{ per tCO}_2, \tag{16}$$

where $\mathrm{SD} = 0.26\,\mathrm{K}$ is the standard deviation of the record. We index by the standard deviation because the range over which the law holds is itself measured in standard deviations.

One number would turn the family into a price, and there are two ways to get it: measure the distance from an element's current state to its threshold in the variable the element is observed in, or calibrate a model of the element's own dynamics in which the stable and unstable states sit a definite distance apart and read the distance off it.

4.2 The endogenous deadline

The decision here is when to stop drawing down the budget. This subsection derives the date at which that happens for the three elements.

4.2.1 From the price to a date

Let the planner choose an abatement rate $a(t) \in [0, 1]$ at cost $c(a)$ per unit of gross emissions E_0, so that $\dot{M} = E_0(1 - a)$. Optimality equates the marginal abatement cost to the shadow value of the remaining budget, $c'(a^*) = \mathrm{SCC}_{\mathrm{threshold}}(\epsilon)$. Suppose the backstop delivers abatement at a constant unit cost $\bar{c}$. The marginal cost is then flat, the first-order condition has no interior solution, and the planner goes all in or does nothing: abatement is full when the price it would avoid exceeds $\bar{c}$ and zero otherwise. The switch occurs where the two are equal, and with $\mathrm{SCC}_{\mathrm{threshold}} = A/\sqrt{\epsilon}$ this inverts in one line,

$$\frac{A}{\sqrt{\epsilon_\infty}} = \bar{c} \quad \Longrightarrow \quad \epsilon_\infty = \left(\frac{A}{\bar{c}}\right)^2, \tag{17}$$

the square coming from inverting the square-root divergence.

Below ϵ_∞ the planner no longer emits, so the budget stops draining and the proximity stops falling. Above it the planner waits and the proximity falls linearly at $\dot{\epsilon} = -E_0/M^*$. Setting $\epsilon(t_c) = \epsilon_\infty$ gives a finite date,

$$t_c = \frac{M^*}{E_0}\big(\epsilon_0 - \epsilon_\infty\big), \tag{18}$$

and this is the sense in which the divergence is not a paralysis. Along the optimal path the price never diverges: it tracks $A/\sqrt{\epsilon(t)}$ until the switch and is capped at $\bar{c}$ thereafter. The formula keeps predicting larger values below ϵ_∞, but optimal policy never enters that region.

Two of the three quantities in (18) are observed and one is not. The proximity ϵ_0 comes from the threshold literature and the ratio M^*/E_0 from the budget and current emissions. The floor is the exception, and it inherits the missing margin twice over: the amplitude carries it as $A \propto \mathcal{L}\, d_{\mathrm{phys}}^{4/3}$, and (17) squares A to invert the square-root divergence, so $\epsilon_\infty \propto \mathcal{L}^2 d_{\mathrm{phys}}^{8/3}$. An elasticity of $8/3$ in the margin against 2 in the loss makes the floor the quantity in this paper that responds most sharply to what has not been measured, and more sharply to the margin than to the damage.

The deadline survives that, and not because the floor is well determined: because it can be dropped. The floor is non-negative and enters (18) with a minus sign, so

$$t_c \;\le\; \frac{M^*}{E_0}\,\epsilon_0 \;=\; \frac{\mu_{\mathrm{now}}}{E_0} \;=\; \frac{(\Delta T^* - \Delta T_0) \times 1{,}000}{\mathrm{TCRE} \times E_0}, \tag{19}$$

where the second equality is the definition of the proximity and the third substitutes (13), in which ΔT^* cancels. The bound is the warming still available divided by the rate at which warming is being spent, and it contains no quantity that has not been measured, though the threshold is measured with considerable uncertainty. At a gross emission rate of $E_0 = 40\,\mathrm{GtCO_2}$ per year (Friedlingstein et al., 2023), the AMOC bound evaluates to 156 years at central inputs.

4.2.2 Which element comes first, and what the backstop cost does not change

Two results follow from the bound, and both use only quantities the literature reports. The first orders the three elements. The second says that the ordering, and the dates themselves, barely depend on how expensive the technology that ends the drawdown turns out to be.

The first result is an ordering between elements. The West Antarctic ice sheet has both the smallest proximity, $\epsilon \approx 0.20$, and a short budget-to-emissions ratio, so its bound is much the nearest, at about 27 years against roughly 129 for the Amazon and 160 for the AMOC. The gap between the ice sheet and the other two is wide enough that no plausible floor would reverse it[4]. The Amazon precedes the AMOC in only 59% of them.

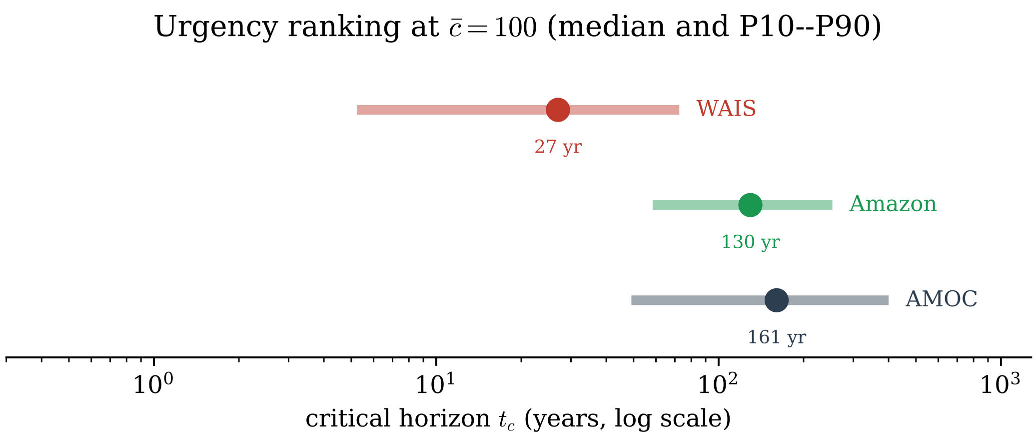

Figure 3: Deadline for the three elements. Median bound (19) (dot) with the $[\mathrm{P10}, \mathrm{P90}]$ interval (bar), from a seeded Monte Carlo over the inputs of Table 3 on a log scale.

The second is that the deadlines barely move with the cost of the technology that ends the drawdown. Raising $\bar{c}$ from \$100 to \$600 leaves the AMOC and Amazon bounds unchanged and shifts the ice sheet's only marginally. The mechanism is the one that produced (19): the backstop cost enters only through the floor $\epsilon_\infty = (A/\bar{c})^2$, and that floor is already negligible, so dividing it by four leaves the date where it was. The timing of optimal action is therefore governed by the geometry of the approach rather than by the cost of the fallback technology, and uncertainty about future abatement costs is not a reason to wait.

4.3 What drives the uncertainty, and what to measure next

Taking logarithms of (15) turns the price into a sum, and at either end of the amplitude's range every coefficient of that sum is a constant. The uncertainty decomposition is then exact rather than local: each input contributes to the variance of the log price in proportion to the square of its elasticity, with no interaction terms and no dependence on the proximity, so the elasticities themselves rank the measurements a planner should want.

What the decomposition runs over has changed, because the reduced form no longer takes the five observables of the opening as its inputs. The budget M^* is determined by the recovery rate through $\lambda = 2\sqrt{M^*\epsilon}$ and no longer varies independently, and the noise σ has

[4]The figures are conditional on no element having crossed already, a condition that removes 25.2% of the West Antarctic draws, since a draw in which the element has already crossed is not one in which it leads but one the price does not describe; including those at $t_c = 0$ would raise the share in which the ice sheet leads from 90% to 91.4%, so the reported figure is the conservative one.

been replaced by the pair formed by that rate and the observed variance. The inputs are the loss $\mathcal{L}$, the realization timescale T, the recovery rate λ, the variance Var, the margin d_{phys}, and the discount rate ρ.

$$\log \mathrm{SCC}_{\mathrm{threshold}} = \mathrm{const} + \log \mathcal{L} + \log \mathsf{f}(\rho T) + \log \Phi_{\mathrm{fold}}(\tilde{\rho}) + \log \eta^* - 2\log\lambda,$$

and substitute the two quantities that carry the observables. The margin and the variance enter through η^*, by (14), as $\log \eta^* = \mathrm{const} + \frac{2}{3}\log d_{\mathrm{phys}} - \frac{1}{3}\log \mathrm{Var}$. They enter a second time through the amplitude, since $\tilde{\rho} = 2\rho\eta^*/\lambda$ carries η^* with it, so each observable reaches the price both directly and through how much the amplitude responds.

Where impatience is small relative to the noise layer the amplitude is close to linear, $\Phi_{\mathrm{fold}}(\tilde{\rho}) \simeq \Phi'_{\mathrm{fold}}(0)\tilde{\rho}$, so $\log \Phi_{\mathrm{fold}}$ contributes $\log\rho + \log\eta^* - \log\lambda$ up to a constant, and

$$\log \mathrm{SCC}_{\mathrm{threshold}} = \mathrm{const} + \log \mathcal{L} + \log \mathsf{f}(\rho T) + \log \rho + \tfrac{4}{3}\log d_{\mathrm{phys}} - \tfrac{2}{3}\log \mathrm{Var} - 3\log\lambda.$$

Every coefficient is a constant, so each input contributes to the variance of the log price in proportion to the square of its elasticity times the variance of its own logarithm, with no interaction terms and no dependence on the proximity.

At the other end, where impatience is large relative to the noise layer, the amplitude grows as the square root of $\tilde{\rho}$ rather than linearly, so its elasticity falls from one to one half and the same substitution gives a second set of coefficients (Appendix A.3).

Table 4: Elasticities of the log tipping price with respect to each input, at the two ends of the amplitude's range. The calibrated elements sit near the small-$\tilde{\rho}$ end, so the left column is the operative one and the right bounds it.

Elasticity of $\log \mathrm{SCC}_{\mathrm{threshold}}$ with respect to	small $\tilde{\rho}$	large $\tilde{\rho}$
Loss at tipping $\mathcal{L}$	1	1
Realization factor $\mathsf{f}(\rho T)$	1	1
Physical margin d_{phys}	$4/3$	1
Observed variance Var	$-2/3$	$-1/2$
Recovery rate λ	-3	$-5/2$
Discount rate ρ, through the amplitude	1	$1/2$

The recovery rate matters most, and by a wide margin: the price moves as its inverse cube, so an estimate good only to a factor of two leaves the price uncertain by a factor of eight. It is also the one of the four that a record can already deliver (Scheffer et al., 2009), which is what puts it first. The other three do not rank the way their elasticities do. How far the element sits from its threshold, in the units it is observed in, moves the price less per unit than the recovery rate, but nobody has measured it at all: Section 2.5.3 puts it above 4.9 standard deviations of the record and nothing caps it from above, which makes it a measurement waiting to be made rather than a parameter to sweep. How long the damage takes to arrive matters for the ice sheets and almost nowhere else, where estimates running from 500 to $13{,}000$ years move the price by a factor of twenty-five, against a span several times narrower for the AMOC. How bad the crossing is moves the price as much as either, and it is the one input this paper does not try to settle. It does not have to: the price is exactly proportional to the loss, so a reader who prefers a different damage estimate rescales the answer and changes nothing else in the calibration.

None of the four input touches what the paper claims: they set the level of the price only. The exponent, the sign of the noise effect and the classification are ratios, and they hold whatever the four turn out to be.

5 Conclusion

This paper has priced one economic object: the marginal value of restraint when a controllable drift pushes a stochastic state toward a boundary whose safety margin collapses as the budget runs out. The first result is that the closing rate of the margin, and nothing else, decides whether the price diverges as the budget empties or stays bounded, so that two systems with identical local dynamics can fall on opposite sides of the split. The second result is that when the control acts on the drift, the price of restraint is higher the weaker the noise driving the system, at every discount rate, as long as the noise remains strong enough that the threshold is what decides the outcome. The second result reverses the standard intuition that more uncertainty raises the price of risk, and it holds because noise and the controller's own push are substitutes in producing a crossing: each is worth less at the margin when the other is abundant. What restraint buys is therefore a later crossing and not a safer one, which is why the threshold component of the price rises with the discount rate where the ordinary Ramsey component falls.

Climate is the canonical application. A quiet element near its threshold is not one with room to spare. When its own fluctuations are weak, what carries it across is the steady pressure we add ourselves, and each unit of restraint bears directly on the outcome. Quiet is the condition under which we are still the ones deciding. Watching an element for signs of agitation is therefore watching for the moment the decision stops being ours.

What a planner obtains from this is a classification, a testable signature, and a deadline. The classification rests on the closing rate of the margin alone, which the structural climate literature reports independently of any damage function, so it can be acted on before the damage debate is settled; on that basis the Atlantic overturning circulation, the Amazon and the two ice sheets are divergent and oscillatory regimes such as ENSO are not. The signature is the cross-partial of the log price in noise and proximity: the ratio of the two prices widens with the remaining budget in both classes, but faster in the divergent one, by a factor running from a few units far from the threshold to two orders of magnitude close to it, so the separation between the classes deepens as the edge is approached.

None of the three elements is inside the window in which the law holds. What the paper prices is therefore the regime they will enter rather than the one they occupy. The deadline survives that gap. Pairing the price with a backstop produces an endogenous budget floor which can only bring the date forward, so the observed proximity and the budget-to-emissions ratio bound it from above without the price, the margin or the recovery rate, and an element outside the window today enters it before its own deadline arrives. The West

Antarctic ice sheet comes first at about 27 years, and the dates barely move with the cost of the backstop: uncertainty about future abatement costs is not a reason to wait. Uncertainty about future abatement costs is not a reason to wait.

The model and the record are not commensurable. The model locates an element in the units of its own equation of motion; a climate record measures the same distance in the units of an instrument. Converting between them takes one number per element, the margin between the current stable state and the unstable one expressed in the element's own observable, and no element has it. Two further measurements would sharpen the level once that one exists, and the elasticities of the closed-form price fix their order. The recovery rate enters with an elasticity of three, which makes it the highest-value target and the one the early-warning literature can already address. The post-tipping timescale enters with an elasticity of one, but across a range wide enough to dominate for the ice sheets.

The analysis faces three limits. The first is the reduction to a single slow variable. It is less restrictive than it appears, since near a generic threshold the remaining modes return to equilibrium faster than the slow direction moves and contribute less than any power of the remaining budget, so the one-dimensional exponent and amplitude carry over intact. Coupling breaks that reduction by introducing several slow directions: for two coupled folds the square-root exponent still survives, but larger networks may concentrate the divergence on a few bottleneck elements whose crossing drags the rest along, and spatially extended systems such as permafrost and boreal forest, where a transition spreads as a front rather than collapsing at a point, plausibly belong to a different class of problem. The second is that the results hold on a window rather than everywhere. They describe the price while the budget drains slowly relative to the system's own recovery and while the stable state still sits inside the noise layer; far from the edge the noise rather than the push does the crossing and the familiar sign returns. The closing stretch of the path, where the boundary itself moves as the state approaches, calls for a dynamic treatment we leave to future work. The third is the standing of one claim about the control channel. A control acting on the volatility rather than the drift does not flip the sign near the threshold, since what sets the sign there is how close the system is and not how the decision-maker acts on it.

The mechanism is not specific to carbon. The same theorem prices any system in which a controllable drift pushes a state toward a collapsing boundary under noise, and its most immediate counterpart is financial stability. Aggregate leverage drifts toward a systemic boundary, and in calm markets it is the steady build-up of exposure rather than volatility that carries the system across, so the value of de-leveraging is highest precisely when markets are quiet. That is the volatility paradox (Brunnermeier and Sannikov, 2014) read as a property of the geometry rather than as a regularity: quiet markets are not safer, they are the ones in which the build-up is the only thing still moving the system. That setting also supplies what the climate one withholds. The distance from a firm's current state to its default boundary is measured, published and available in panel, so the ratio this paper cannot form for a climate element can be formed there. Both results become testable: whether the price of prudence rises as leverage builds, and at what rate the margin closes, which is what decides whether a systemic price diverges at all. What a price near a point of no return must reflect is the rate at which the control closes the margin, not the size of the noise around it.

References

Anthoff, D., F. Estrada, and R. S. Tol (2016). Shutting down the thermohaline circulation. *American Economic Review 106*(5), 602–606.

Armstrong McKay, D. I., A. Staal, J. F. Abrams, R. Winkelmann, B. Sakschewski, S. Loriani, I. Fetzer, S. E. Cornell, J. Rockström, and T. M. Lenton (2022). Exceeding 1.5° c global warming could trigger multiple climate tipping points.

Ashwin, P., S. Wieczorek, R. Vitolo, and P. Cox (2012). Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 1166–1184.

Boulton, C. A., T. M. Lenton, and N. Boers (2022). Pronounced loss of amazon rainforest resilience since the early 2000s. *Nature Climate Change 12*(3), 271–278.

Brunnermeier, M. K. and Y. Sannikov (2014). A macroeconomic model with a financial sector. *American Economic Review 104*(2), 379–421.

Caesar, L., S. Rahmstorf, A. Robinson, G. Feulner, and V. Saba (2018). Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature 556*(7700), 191–196.

Cai, Y. and T. S. Lontzek (2019). The social cost of carbon with economic and climate risks. *Journal of Political Economy 127*(6), 2684–2734.

Cessi, P. (1994). A simple box model of stochastically forced thermohaline flow. *Journal of physical oceanography 24*(9), 1911–1920.

Dietz, S., J. Rising, T. Stoerk, and G. Wagner (2021). Economic impacts of tipping points in the climate system. *Proceedings of the National Academy of Sciences 118*(34), e2103081118.

Ditlevsen, P. and S. Ditlevsen (2023). Warning of a forthcoming collapse of the atlantic meridional overturning circulation. *Nature communications 14*(1), 4254.

Fillon, R. and C. Guivarch (2023). The need for regulation of climate subsystems. Technical report.

Fillon, R., C. Guivarch, and N. Taconet (2023). Optimal climate policy under tipping risk and temporal risk aversion. *Journal of Environmental Economics and Management 121*, 102850.

Friedlingstein, P., M. O'sullivan, M. W. Jones, R. M. Andrew, D. C. Bakker, J. Hauck, P. Landschützer, C. Le Quéré, I. T. Luijkx, G. P. Peters, et al. (2023). Global carbon budget 2023. *Earth System Science Data 15*(12), 5301–5369.

Haigh, M. and P. R. Holland (2024). Decadal variability of ice-shelf melting in the amundsen sea driven by sea-ice freshwater fluxes. *Geophysical Research Letters 51*(9), e2024GL108406.

IPCC (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781009157896.

Jenkins, A., D. Shoosmith, P. Dutrieux, S. Jacobs, T. W. Kim, S. H. Lee, H. K. Ha, and S. Stammerjohn (2018). West antarctic ice sheet retreat in the amundsen sea driven by decadal oceanic variability. *Nature Geoscience 11*(10), 733–738.

Jin, F.-F. (1997). An equatorial ocean recharge paradigm for enso. part i: Conceptual model. *Journal of the atmospheric sciences 54*(7), 811–829.

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

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

Pindyck, R. S. (2000). Irreversibilities and the timing of environmental policy. *Resource and energy economics 22*(3), 233–259.

Pindyck, R. S. and N. Wang (2013). The economic and policy consequences of catastrophes. *American Economic Journal: Economic Policy 5*(4), 306–339.

Robinson, A., R. Calov, and A. Ganopolski (2012). Multistability and critical thresholds of the greenland ice sheet. *Nature climate change 2*(6), 429–432.

Scheffer, M., J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara (2009). Early-warning signals for critical transitions. *Nature 461*(7260), 53–59.

Schoof, C. (2007). Ice sheet grounding line dynamics: Steady states, stability, and hysteresis. *Journal of Geophysical Research: Earth Surface 112*(F3).

Staal, A., I. Fetzer, L. Wang-Erlandsson, J. H. Bosmans, S. C. Dekker, E. H. van Nes, J. Rockström, and O. A. Tuinenburg (2020). Hysteresis of tropical forests in the 21st century. *Nature communications 11*(1), 4978.

Van Den Bremer, T., C. Hambel, and F. van der Ploeg (2026). Three reasons to price carbon under uncertainty: Accuracy of simple rules. *Quantitative Economics 17*(3), 858–894.

van den Bremer, T. S. and F. van der Ploeg (2021). The risk-adjusted carbon price. *American Economic Review 111*(9), 2782–2810.

van der Ploeg, F. and A. de Zeeuw (2019). Pricing carbon and adjusting capital to fend off climate catastrophes. *Environmental and Resource Economics 72*(1), 29–50.

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

A Appendix

A.1 The reduced problem, step by step

Section 2.1 compresses into a page the passage from a three-state growth problem to the one-dimensional boundary-value problem (4) that carries every result in the paper. We present the mechanism: why logarithmic utility separates the value function additively (A.1.1), how the Hamilton–Jacobi–Bellman equation reduces to (2) (A.1.2), how the price is rewritten in the remaining-budget fraction (A.1.3), and where the crossing equation sits inside the reduced problem (A.1.4).

A.1.1 Why logarithmic utility separates the value function additively

The body posits $V(K, M, X) = \frac{1}{\rho} \log K + v(M, X) + C_0$ and verifies it; the form can also be derived. Every term in $\dot{K} = AK - c - C(a)K$ is homogeneous of degree one in (K, c), so scaling initial capital by λ scales the feasible consumption set by λ, while the climate block is untouched since neither $(1-a)E_0$ nor $f(X; M)$ involves K. Optimality is therefore preserved by the rescaling, and

$$V(\lambda K, M, X) = \max \int_0^\infty e^{-\rho t} \log(\lambda c_t)\, dt = V(K, M, X) + \frac{\log \lambda}{\rho}, \tag{20}$$

because the constant $\log \lambda$ integrates to $\log \lambda/\rho$ against the discount factor. Setting $\lambda = K$ and evaluating at unit capital gives the separation directly,

$$V(K, M, X) = \frac{1}{\rho} \log K + \underbrace{V(1, M, X)}_{\equiv\, v(M, X) + C_0}.$$

The coefficient $1/\rho$ is not a free constant to be matched: it is the discounted value of a permanent unit flow. What logarithmic utility contributes is that the wealth effect enters inside the utility function additively, $\log(\lambda c) = \log \lambda + \log c$. Under a general power utility $u(c) = c^{1-\gamma}/(1-\gamma)$ the same logic returns $V(\lambda K, M, X) = \lambda^{1-\gamma} V(K, M, X)$, hence the multiplicative form $V = \frac{K^{1-\gamma}}{1-\gamma} G(M, X)$.

The cross-derivative marks the two cases apart without being what separates them. Adding the wealth term leaves $V_K = 1/(\rho K)$, free of the climate state, so $\partial^2 V/\partial K \partial M = 0$; multiplying it in leaves $V_K = K^{-\gamma} G(M, X)$, which the climate state enters. That difference does not reach the price, because the price is the ratio $-V_M/V_K$ and the factor carrying the climate state cancels from it under either form, leaving $\rho K(-v_M)$ in one case and $-K\, \partial_M \log G/(1-\gamma)$ in the other. What Section 3.1.2 uses is a property of the reduced climate equation instead: its no-tipping solution factors out, leaving a problem for the threshold correction alone.

The same homogeneity explains why the crossing penalty can be a constant $-\mathcal{L}$ independent of K. Suppose tipping permanently destroys a fraction D of the capital stock or, equivalently, scales the whole subsequent consumption path by $1 - D$. By (20) the value falls by exactly $\frac{1}{\rho} \log(1 - D)$, so $\mathcal{L} = -\frac{1}{\rho} \log(1 - D)$, a number and not a function of wealth. Under power utility the same damage would cost $\propto K^{1-\gamma}$ and the boundary condition would carry the capital stock back into the climate block.

A.1.2 The Hamilton–Jacobi–Bellman equation, step by step

Step 0: primitives. The states are the capital stock K, the accumulated emission stock M, and the climate indicator X; the controls are consumption c and the abatement rate $a \in [0,1]$. Their laws of motion are

$$dK = \big(AK - c - C(a)K\big)dt, \qquad dM = (1-a)E_0\,dt, \qquad dX = f(X;M)\,dt + \sigma\,dW_t,$$

the last of which is (1). Only X carries a diffusion. The planner maximises $\mathbb{E}\int_0^\infty e^{-\rho t}\log c_t\,dt$ with the one-time loss $\mathcal{L}$ at the crossing date τ.

Step 1: the dynamic programming principle. Over a short interval, the planner consumes and then holds the same problem from the new state,

$$V(K,M,X) = \max_{c,a}\ \mathbb{E}\Big[\log c\,dt + e^{-\rho\,dt}\,V(K+dK,\ M+dM,\ X+dX)\Big] + o(dt).$$

Step 2: expand the continuation value. By Itô's lemma,

$$dV = V_K\,dK + V_M\,dM + V_X\,dX + \tfrac{1}{2}V_{XX}(dX)^2.$$

No other second-order term appears. The stocks K and M have finite variation, so $(dK)^2$ and $(dM)^2$ are $O(dt^2)$ and the cross terms $dK\,dX$ and $dM\,dX$ are $O(dt^{3/2})$; all vanish in the limit. Taking expectations and using $\mathbb{E}[dX] = f\,dt$ and $\mathbb{E}[(dX)^2] = \sigma^2\,dt + O(dt^{3/2})$,

$$\mathbb{E}[dV] = \Big(V_K\dot{K} + V_M\dot{M} + f\,V_X + \tfrac{\sigma^2}{2}V_{XX}\Big)dt + o(dt).$$

Step 3: pass to the limit. Expanding $e^{-\rho\,dt} = 1 - \rho\,dt + O(dt^2)$, substituting, cancelling V on both sides, dividing by dt and letting $dt \to 0$ gives the Hamilton–Jacobi–Bellman equation of the body,

$$\rho V = \max_{c,\ a\in[0,1]}\Big\{\log c + V_K\big(AK - c - C(a)K\big) + V_M\,(1-a)E_0 + f(X;M)\,V_X + \frac{\sigma^2}{2}V_{XX}\Big\},$$

the emission term being $V_M\dot{M}$ with $\dot{M} = (1-a)E_0$, since the stock is carried absolutely. The budget to threshold enters when the problem is rewritten in the fraction $\epsilon = (M^* - M)/M^*$.

Step 4: substitute the separation and the consumption rule. With $V = \frac{1}{\rho}\log K + v(M,X) + C_0$ one has $V_K = 1/(\rho K)$, so the first-order condition in consumption, $1/c^* = V_K$, gives $c^* = \rho K$, while V_M, V_X, V_{XX} are simply v_M, v_X, v_{XX}. Term by term:

$$\rho V = \log K + \rho v + \rho C_0, \qquad \log c^* = \log\rho + \log K, \qquad V_K\big(AK - c^* - C(a)K\big) = \frac{A - \rho - C(a)}{\rho}.$$

The third identity is where the capital stock leaves the problem: the linear return AK divided by the marginal utility $1/(\rho K)$ is free of K. This is the sense in which AK and log are complementary: log makes K appear additively as $\log K$ on both sides, AK makes everything else independent of it.

Step 5: cancel and absorb. The term $\log K$ appears on both sides and cancels. Choosing C_0 to absorb $\log\rho + (A-\rho)/\rho$ leaves

$$\rho v = \max_{a\in[0,1]}\Big\{-\frac{C(a)}{\rho} + v_M\,(1-a)E_0\Big\} + f\,v_X + \frac{\sigma^2}{2}v_{XX},$$

which rearranges to (2), with the crossing penalty entering only through $v(M, x_-(\mu)) = -\mathcal{L}$.

A.1.3 From the stock derivative to the budget-fraction derivative

Equation (3) rewrites the price in the remaining-budget fraction. Three things happen at once, and separating them removes the apparent jump.

First, the price loads only on v. Since $\rho/\log K$ and C_0 do not depend on the stock, $\Pi = -\partial_M V = -\partial_M v$. Second, v depends on the stock only through φ. Writing,

$$v(M, X) = -\mathcal{L}\,\varphi(X;\epsilon) + g(M) + w(M, X),$$

the tipping channel is carried entirely by φ and the remaining terms are the ordinary Ramsey channel. Third, the change of variable. With $\mu = M^* - M$ and $\epsilon = \mu/M^*$,

$$\frac{\partial\epsilon}{\partial M} = -\frac{1}{M^*} \qquad\Longrightarrow\qquad \partial_M v = -\mathcal{L}\,\partial_\epsilon\varphi\cdot\Big(-\frac{1}{M^*}\Big) = \frac{\mathcal{L}}{M^*}\,\partial_\epsilon\varphi.$$

Since a larger remaining budget places the stable state further from the boundary, $\partial_\epsilon\varphi < 0$; hence $\partial_M v < 0$, emitting reduces welfare, thus yields (3)

$$\Pi = -\partial_M v = \frac{\mathcal{L}}{M^*}\Big|\frac{\partial\varphi}{\partial\epsilon}\Big|,$$

The factor $1/M^*$ is a Jacobian and carries the units: ϵ is dimensionless, so $\partial_\epsilon\varphi$ is a sensitivity per unit of budget fraction, and dividing by M^* converts it into a sensitivity per unit of budget. It is also the origin of the $\sqrt{M^*}$ in the closed form. Anticipating Appendix A.5.1, with $\varphi(x^*) = 1 - \frac{2\Phi_{\text{fold}}(\tilde\rho)}{\sigma^{2/3}}\sqrt{\mu} + O(\mu)$ and $\mu = M^*\epsilon$,

$$\partial_\epsilon\varphi = -\frac{2\Phi_{\text{fold}}(\tilde\rho)}{\sigma^{2/3}}\cdot\frac{\sqrt{M^*}}{2\sqrt{\epsilon}} \qquad\Longrightarrow\qquad \Pi = \frac{\mathcal{L}}{M^*}\cdot\frac{\Phi_{\text{fold}}(\tilde\rho)\sqrt{M^*}}{\sigma^{2/3}\sqrt{\epsilon}} = \frac{\mathcal{L}\,\Phi_{\text{fold}}(\tilde\rho)}{\sigma^{2/3}\sqrt{M^*}}\,\epsilon^{-1/2},$$

which is (9) with every constant accounted for.

$\Pi = -\partial_M V$ is a shadow price in units of utility. The carbon price in consumption units divides it by the marginal utility of consumption, $\text{SCC} = -V_M/V_K = \rho K\,(-v_M)$. The body works with Π throughout because the conversion factor ρK multiplies both channels identically and therefore affects neither the divergence exponent, nor the sign of the noise effect, nor the cross-partial.

A.1.4 The crossing equation

Fix the budget and let $\varphi(x) = \mathbb{E}_x[e^{-\rho\tau}]$ with $\tau = \inf\{t : X_t \le x_-\}$. The Feynman–Kac representation gives (4) directly,

$$\tfrac{\sigma^2}{2}\varphi'' + f\varphi' = \rho\varphi, \qquad x\in(x_-,\infty),$$

with $\varphi(x_-) = 1$ because crossing is immediate there, and $\varphi(\infty) = 0$ because crossing is pushed into an infinitely discounted future. Read as an asset price, $e^{-\rho t}\varphi(X_t)$ is a martingale up to τ: the claim pays no dividend before crossing, so its discounted price does not drift.

The same operator governs the reduced problem (2), and that is what connects the two. Write $\mathcal{A} = \frac{\sigma^2}{2}\partial_{XX} + f\partial_X - \rho$, so that the crossing equation is $\mathcal{A}\varphi = 0$ and the reduced problem is $\mathcal{A}v = -\max_a\{\cdot\}$ with $v(M, x_-(\mu)) = -\mathcal{L}$ at the boundary. Where the source term vanishes, the solution is exactly $v = -\mathcal{L}\varphi$: it satisfies the equation because $\mathcal{A}\varphi = 0$, and the boundary condition because $\varphi(x_-) = 1$. Elsewhere the abatement cost adds a forced part, so φ is the homogeneous half of v and everything else is what the cost drives.

A.2 The two scales in closed form (H2)

Hypothesis (H2) supposes that the noise level and the discount rate enter the price only through a single ratio. This appendix shows that for the geometries we treat this is a consequence rather than an assumption: measuring the state in noise-layer widths removes the noise from the equation everywhere except one place, where it appears dividing the discount rate, and that combination is $\tilde{\rho}$. Set $x = \ell\eta$ in (4) at the threshold, where the drift is $-c_k x^k$. The substitution reaches both the derivatives and the coefficient: each derivative in the state costs a factor $1/\ell$, and x^k becomes $\ell^k\eta^k$. The three terms therefore pick up different powers,

$$\frac{\sigma^2}{2\,\ell^2}\,\varphi_{\eta\eta} - c_k\,\ell^{\,k-1}\,\eta^k\varphi_\eta = \rho\,\varphi,$$

the drift carrying ℓ^k from the coefficient against ℓ^{-1} from the derivative. Dividing through by that drift size $\ell^{\,k-1}$ leaves

$$\frac{\sigma^2}{2\,\ell^{\,k+1}}\,\varphi_{\eta\eta} - c_k\,\eta^k\varphi_\eta = \frac{\rho}{\ell^{\,k-1}}\,\varphi,$$

The choice $\ell^{k+1} = \sigma^2$ that fixes the noise layer is exactly the one making the first coefficient $\frac{1}{2}$, a constant with no parameter left in it. The equation becomes

$$\tfrac{1}{2}\,\varphi_{\eta\eta} - c_k\,\eta^k\varphi_\eta = \tilde{\rho}\,\varphi, \qquad \tilde{\rho} = \frac{\rho}{m(\sigma)}, \qquad m(\sigma) = \ell(\sigma)^{k-1},$$

in which neither the noise level nor the discount rate appears alone. That is (H2), produced rather than posited, and both exponents follow from what has already been fixed. The layer was defined by $\ell^{k+1} = \sigma^2$, which solves to $\ell = \sigma^a$ with $a = 2/(k+1)$; the second scale is $m = \ell^{\,k-1}$, so $b = (k-1)\,a$ and

$$a = \frac{2}{k+1}, \qquad b = \frac{2(k-1)}{k+1}.$$

Both are positive for every $k \geq 2$, so the requirement $b > 0$ follows from two equilibrium branches merging rather than being an extra restriction.

The same two exponents give the elasticity of the price to the noise, which the body quotes. The noise reaches the price only through the steepness $\kappa = \Phi(\tilde{\rho})/\ell(\sigma)$, and it reaches it twice: explicitly through $\ell = \sigma^a$, and inside the amplitude through $\tilde{\rho} = \rho/m(\sigma) = \rho\,\sigma^{-b}$. How much the second channel counts depends on where the amplitude is read.

Where impatience is small the amplitude is close to linear, so $\Phi \simeq \Phi'(0)\rho\,\sigma^{-b}$, the steepness is $\Phi'(0)\,\rho\,\sigma^{-(a+b)}$, and the elasticity is $-(a+b) = -2k/(k+1)$, which returns $-4/3$ at $k = 2$ and $-3/2$ at $k = 3$. Where impatience is large the amplitude behaves as the square root of $\tilde{\rho}$ instead, which is free of k because that limit is set by diffusion against discounting with the drift negligible, and the drift is the only place k enters. Then $\Phi \propto \sqrt{\rho}\,\sigma^{-b/2}$, so the second channel enters at half weight and

$$-\Big(a + \frac{b}{2}\Big) = -\Big(\frac{2}{k+1} + \frac{k-1}{k+1}\Big) = -1 \qquad \text{for every } k.$$

A.3 Verifying the amplitude hypothesis (H3)

(H3) requires that the amplitude is positive and increasing in the rescaled discount rate: the reversal rests on that single property, so we settle it in closed form rather than numerically. The route removes the first-derivative term from the pricing equation, measures the state in units of the noise layer, identifies the one number the price reads off the resulting equation, evaluates it at zero discounting, and signs its derivative. We carry it through for the fold, then show in Section A.3.5 that it holds at every local order without modification.

A.3.1 Removing the first-derivative term

The first-derivative term of (4) is the awkward one, because its coefficient is the drift and therefore varies with the state. Changing the dependent variable removes it, as a shift of variable removes the linear term of a quadratic. Write $\varphi = Gu$ and ask what G must be for the terms in u' to cancel. Their coefficient is $\sigma^2 G' + fG$, so the requirement is $G'/G = -f/\sigma^2$, that is $G = e^{-B/\sigma^2}$ with $B' = f$, the potential whose slope is the drift. The factor is solved for rather than guessed. Substituting and dividing through by it leaves

$$u'' = \frac{2}{\sigma^2}\,W(x;\mu)\,u, \qquad W = \rho + \frac{f'}{2} + \frac{f^2}{2\sigma^2}, \tag{21}$$

which does not depend on the geometry, so Section A.3.5 can run it at any local order; for the fold, $f = \mu - x^2$ gives $f'/2 = -x$ and $f^2/(2\sigma^2) = (\mu - x^2)^2/(2\sigma^2)$.

One feature of W matters for what follows: the discount rate enters it as an additive constant where the two geometric terms carry powers of the state. That is what makes the three terms scale differently under the rescaling of the next part, and it is the only channel through which impatience reaches the amplitude.

A.3.2 Measuring the state in noise-layer units

The rescaling does two things. The first is that it strips the noise out of the equation: measured in the right units, the two terms that came from the geometry lose their dependence on σ entirely, and the discount rate survives only in the combination $\tilde{\rho}$. The second is that the equation it leaves does not determine its own solution, only fixing it up to a multiplicative constant, so one has to ask what the price can read off an object known only up to scale.

The answer is a single number, the logarithmic slope at the origin, and that number is the amplitude.

Equation (21) still carries three parameters, and the rescaling removes two. Set $\mu = 0$, which places the problem at the threshold where the amplitude is defined, and measure the state in noise-layer units by writing $x = \sigma^{2/3}\eta$. The left-hand side becomes $\sigma^{-4/3}u_{\eta\eta}$, so the equation reads $u_{\eta\eta} = (2W/\sigma^{2/3})u$, and the three terms of W scale differently, which is where the rescaling earns its keep: the discount rate is unchanged, the mean-reversion term becomes $\sigma^{2/3}\eta$ and the quartic term $\sigma^{2/3}\eta^4/2$, so dividing through leaves

$$u''(\eta) = \left(2\tilde{\rho} - 2\eta + \eta^4\right) u(\eta), \qquad \tilde{\rho} = \frac{\rho}{\sigma^{2/3}}. \tag{22}$$

The two terms that came from the geometry have lost their dependence on the noise entirely, and the discount rate survives only in the combination $\tilde{\rho}$. That is (H2) made explicit.

The equation does not pin down u itself, only its shape: the solutions that decay away from the boundary are all multiples of one another, and nothing selects the multiple. A single value of u therefore means nothing on its own, and the price does not use one. It uses a comparison of two, in which the unknown multiple cancels. The two are the crossing value at the stable state and at the boundary, which in rescaled coordinates sit symmetrically at $\eta = \pm\eta^*$, so for small η^*,

$$\frac{u(\eta^*)}{u(-\eta^*)} = \frac{u(0) + \eta^* u'(0) + O(\eta^{*2})}{u(0) - \eta^* u'(0) + O(\eta^{*2})} = 1 + \frac{2u'(0)}{u(0)}\,\eta^* + O(\eta^{*2}),$$

The price differentiates the shortfall below certainty rather than the ratio itself, so subtract from one:

$$1 - \frac{u(\eta^*)}{u(-\eta^*)} = -\frac{2u'(0)}{u(0)}\,\eta^* + O(\eta^{*2}).$$

The shortfall is positive, the crossing value being lower at the stable state than at the boundary, so the quotient it is built from is negative and the sign has to be carried into the definition. The inner equation (22) leaves $\tilde{\rho}$ as its only parameter, so that quotient depends on $\tilde{\rho}$ and on nothing else. We call it the amplitude

$$\Phi_{\text{fold}}(\tilde{\rho}) \;\equiv\; -\frac{u'(0;\tilde{\rho})}{u(0;\tilde{\rho})}. \tag{23}$$

The remaining factor of two cancels against the one in $\eta^* = d/2\ell$, leaving the shortfall as $\Phi_{\text{fold}}\, d/\ell$, which is the $\kappa\, d$ of Section 2.4 with $\kappa = \Phi_{\text{fold}}/\ell$. The origin is the right point to evaluate it at, being the merged equilibrium about which the two turning points sit symmetrically as the budget runs out.

A.3.3 The amplitude at zero discounting

The value $\Phi_{\text{fold}}(0) = 0$ follows without solving anything. At $\rho = 0$ the crossing value is $\mathbb{E}_x[e^{-0\cdot\tau}] = 1$ for every starting point, since the boundary is reached in finite time with probability one at any fixed budget. The crossing value is then constant in the state and in the budget, every derivative of it vanishes, and the amplitude, being a rescaled version of one of those derivatives, vanishes with them.

The economic reading is direct. A planner who does not discount is indifferent between an early tipping and a late one, so restraint, which buys only delay, is worth nothing to him. The price of restraint is the price of delay, and it falls to zero when delay is worth nothing. That is the origin of the comparative static in the discount rate reported in Section 3.3.

A.3.4 Signing the derivative without solving

The amplitude has no closed form, so the sign of its derivative has to be settled without computing the function. We have an equation the solution satisfies; we differentiate that equation in the parameter, which gives a second one describing how the solution moves when the parameter changes; and we combine the two so that everything unknown cancels. What survives is an identity from which the sign follows.

Write u for the decaying solution of (22) and $\dot{u}$ for how it shifts when the discount rate is raised. Two equations are then available, the original one and the one obtained by differentiating it in $\tilde{\rho}$,

$$u'' = Vu, \qquad \dot{u}'' = V\dot{u} + 2u,$$

the extra $2u$ coming from the discount rate entering the potential as an additive constant. Neither equation can be solved, but the pair can be combined so that the unknown potential drops out. Multiplying the first by $\dot{u}$, the second by u and subtracting, everything involving V cancels and the left-hand side is a derivative,

$$\big(u'\dot{u} - \dot{u}'u\big)' = -2u^2. \tag{24}$$

The identity is useful because the expression on its left is exactly what the amplitude's derivative is built from. Indeed $\Phi_{\text{fold}} = -u'/u$ at the origin gives $\Phi'_{\text{fold}} = (u'\dot{u} - \dot{u}'u)/u^2$ there. Integrating outward settles the sign, the expression vanishing far away because the decaying solution falls faster than any power, so

$$\Phi'_{\text{fold}}(\tilde{\rho}) = \frac{2}{u(0;\tilde{\rho})^2}\int_0^\infty u(\eta;\tilde{\rho})^2\,d\eta \;>\; 0, \tag{25}$$

a squared quantity integrated and divided by another is positive whatever the discount rate.

The same three gestures give the concavity, with one extra step at the end. Differentiating the pair a second time in $\tilde{\rho}$ leaves $\ddot{u}'' = V\ddot{u} + 4\dot{u}$ alongside the original, and the same subtraction gives $(u'\ddot{u} - \ddot{u}'u)' = -4u\dot{u}$. The right-hand side is no longer a square, but (24) integrated outward turns $\dot{u}/u$ into an expression that, carried in and integrated by parts, leaves

$$\Phi''_{\text{fold}}(\tilde{\rho}) = -\frac{2}{u(0)^2}\int_0^\infty u(\eta)^2\left(\frac{d}{d\eta}\frac{\dot{u}}{u}\right)^2 d\eta \;<\; 0, \tag{26}$$

a square again, built from how the slope itself responds to the discount rate. Impatience therefore steepens the crossing profile by less at each further increment.

Those two results are (H3). Starting from $\Phi_{\text{fold}}(0) = 0$ and integrating a strictly positive derivative gives $\Phi_{\text{fold}} > 0$ and $\Phi'_{\text{fold}} > 0$ for every $\tilde{\rho} > 0$, so the amplitude is positive and strictly increasing throughout.

A.3.5 The same argument at every local order

The argument just given never used the shape of the potential, so it never used k. It used two properties of it: that the discount rate enters as an additive constant, which is what produced the $2u$ on differentiating, and that the potential cancels in the subtraction, which is what allowed the identity to be written without solving anything. Both hold at every local order. At order k the interior problem is

$$u'' = \big(2\tilde{\rho} - k\,\eta^{k-1} + \eta^{2k}\big)u, \tag{27}$$

which is (22) at $k = 2$. Its two state-dependent terms both descend from the drift $-x^k$, one through $f'/2$ and the other through $f^2/2\sigma^2$, so a single exponent governs both while the discount rate sits apart from them as before. The decaying solution at zero discounting is $u_0 = e^{-\eta^{k+1}/(k+1)}$.

Each step then goes through unchanged. Differentiating in $\tilde{\rho}$ leaves $\dot{u}'' = V\dot{u} + 2u$, with the same $2u$ and for the same reason; the same subtraction gives (24); and u still falls faster than any power, so $u'\dot{u} - \dot{u}'u$ still vanishes far from the origin. So (25) holds with Φ_k in place of Φ_{fold}, and differentiating a second time gives $\Phi_k'' < 0$. Positivity, monotonicity and concavity hold at every k, which is (H3) in full generality. That leaves (H2) as the only hypothesis whose form differs across the rows of Table 1.

A.4 The error the additive split leaves

Section 3.1.2 states the split around (10). Here we derive it and bound its error. Section 2.1 separates the value function as $V(K, M, X) = \rho^{-1}\log K + v(M, X) + C_0$, from which the price reads $\text{SCC} = \rho K\, v_\epsilon / M^*$, so bounding an error on the price means bounding one on v_ϵ. Write $v = -\mathcal{L}\,\varphi(X;\epsilon) + g(\epsilon) + w(X, \epsilon)$, with φ the homogeneous solution of (4). Where the planner waits, the abatement cost drops out and the reduced problem is $\mathcal{A}v = v_\epsilon E_0/M^*$, with $\mathcal{A} = \frac{\sigma^2}{2}\partial_{XX} + f\partial_X - \rho$ and $v = -\mathcal{L}$ at the boundary.

Substituting sorts the terms by whether they carry the climate state. Those that do not give an equation in ϵ alone: g has no X-derivative, so $\mathcal{A}g = -\rho g$, and collecting leaves $-\rho g = (E_0/M^*)g'$, which is the precise sense in which the Ramsey channel does not see the climate state. Those that do vanish for φ itself, by $\mathcal{A}\varphi = 0$, except where φ is differentiated in the budget. What survives is carried entirely by w,

$$\mathcal{A}w - \frac{E_0}{M^*}\,w_\epsilon = r, \qquad r = -\frac{E_0}{M^*}\,\mathcal{L}\,\varphi_\epsilon, \tag{28}$$

and the forcing names the omission: φ_ϵ is the rate at which the crossing value itself changes as the budget drains, which a calculation run at a frozen budget leaves out, and it is the one term the decomposition does not absorb.

Bounding it is a matter of discounting. The operator on the left discounts at rate ρ and the w_ϵ term on the left adds no zeroth-order contribution, so a bounded forcing produces a response no larger than the forcing divided by ρ, giving $|w| \le |r|/\rho$. The price needs w_ϵ rather than w. Differentiating (28) in ϵ returns the same operator acting on w_ϵ, with r_ϵ in place of r and one further term, $(\partial_\epsilon f)\,w_X$, because the drift and the boundary both move with the budget. Since r_ϵ carries $\varphi_{\epsilon\epsilon}$ where r carried φ_ϵ, and $\varphi_\epsilon \propto \epsilon^{p-1}$ by (6) makes $\varphi_{\epsilon\epsilon}/\varphi_\epsilon$ equal to $(p-1)/\epsilon$,

$$\frac{|w_\epsilon|}{\mathcal{L}\,|\varphi_\epsilon|} = O\Big(\frac{E_0}{\rho M^*\,\epsilon}\Big), \tag{29}$$

so the split is usable whenever $\epsilon \gg E_0/(\rho M^*)$. At $p = 1$ the ratio vanishes and the leading error with it, so the bounded class satisfies the condition automatically. Abatement only relaxes the bound further: it multiplies that rate by $1 - a \le 1$ whatever rule sets a, so the estimate holds a fortiori, and at full abatement the budget stops moving and the split is exact.

A.5 The cross-partial signature

Section 3.2 states that the two classes differ in the cross-partial of the log price, that the difference follows from the closing rate alone, and that its size on the observable range is read off the numerical solution. This appendix derives the price and gives the cross-partial in closed form.

A.5.1 From the amplitude to the price

The amplitude is a slope of the rescaled solution and the price is a derivative in the budget. One step joins them: undo the change of variable of Section A.3.1 and read the crossing value at the two points that matter.

At budget $\mu > 0$ the fold has its stable state and its boundary at $\pm\sqrt{\mu}$, which in rescaled coordinates sit symmetrically at $\pm\eta^*$. Writing $\varphi = e^{-B/\sigma^2}u$ with $B(x) = \mu x - \frac{1}{3}x^3$ and imposing $\varphi(x_-) = 1$,

$$\varphi(x^*;\mu) = e^{-\frac{4}{3}\eta^{*3}}\ \frac{u(\eta^*)}{u(-\eta^*)}.$$

The first factor differs from one by a third power of η^*, so it does not reach either of the two orders below. What remains is the ratio of Section A.3.2, read across the whole span from the boundary to the stable state. That span is $2\eta^*$ wide, so the shortfall below certainty is $2\Phi_{\text{fold}}\eta^*$: the amplitude times the margin, measured in noise-layer widths.

Because the two points sit symmetrically about the origin, the next term is settled by the first and brings in no further property of u. It is half the leading shortfall with the sign reversed, so $1 - \varphi(x^*) = 2\Phi_{\text{fold}}\eta^*\big(1 - \Phi_{\text{fold}}\eta^* + O(\eta^{*2})\big)$. Differentiating in the budget and converting to ϵ reproduces (9) with its leading correction explicit,

$$\mathrm{SCC}_{\text{threshold}}(\epsilon) = \mathcal{L}\cdot\mathsf{f}(\rho T)\cdot\frac{\Phi_{\text{fold}}(\tilde\rho)}{\sigma^{2/3}\sqrt{M^*}}\cdot\epsilon^{-1/2}\big(1+K_1\sqrt{\epsilon}+O(\epsilon)\big), \qquad K_1 = -\frac{2\,\Phi_{\text{fold}}(\tilde\rho)\sqrt{M^*}}{\sigma^{2/3}} < 0. \tag{30}$$

The correction is negative, so the leading law overstates the price at any finite distance from the edge. It is proportional to the steepness Φ_{fold}/ℓ, which is what puts the cross-partial in closed form below.

A.5.2 The coefficient in closed form

The correction to the leading price is cleanest in twice the ratio of Section 2.4, $z = 2\eta^* = d(\mu)/\ell(\sigma)$, in which (30) reads $-\Phi_{\text{fold}}(\tilde\rho)\,z$. With $d(\mu) = c\,\mu^p$ it goes as ϵ^p. For any geometry satisfying the hypotheses, the mixed derivative is then

$$\frac{\partial^2 \log \mathrm{SCC}_{\text{threshold}}}{\partial \log\sigma\,\partial\log\epsilon} = p\,z\left[\,a\,\Phi_{\text{fold}}(\tilde\rho) + b\,\tilde\rho\,\Phi'_{\text{fold}}(\tilde\rho)\right], \tag{31}$$

with p the closing rate of (H1) and a, b the two scale exponents of (H2). The bracket is the one that signs the reversal, positive by (H3), and it appears here for the same reason it appears there: the coefficient of the correction is proportional to the steepness, whose response to noise (8) already signs. Both geometries treated here have $k = 2$ and therefore $a = b = \frac{2}{3}$; the fold, with $p = \frac{1}{2}$, gives $\frac{2}{3}\sqrt{M^*\epsilon}\,[\cdot]/\sigma^{2/3}$, which is (30) differentiated, and the transcritical, with $p = 1$, gives $\frac{2}{3}M^*\epsilon\,[\cdot]/\sigma^{2/3}$. The two consequences of Theorem 1 are therefore one object seen at two orders: the bracket that signs the reversal, carried by the power that fixes the divergence.

A.5.3 What the closed form settles, and what it does not

Equation (31) is exact, and it is not what governs the range on which the signature is measured. It fixes the structure; the size on the observable range comes from the numerical solution. The reason is that the next term carries the opposite sign and a larger coefficient, so the mixed derivative turns negative once the stable state sits far enough out toward the edge of the layer. Fitting the numerical solution locates that turn at

$$\eta^*_{\text{flip}} \approx 0.71\,\tilde\rho^{\,1.04}.$$

The turn is not a fixed threshold: $\tilde\rho = 2\rho\eta^*/\lambda$ carries η^* with it, so the turn rises as the element does.

Whether an element clears it is therefore a question about the ratio of the two, and that ratio has an unusual property for a quantity built on the margin: it does not depend on the missing conversion. It varies as $\eta^{*\,1-1.04}$, an exponent near enough to zero that it moves by three percent across the whole plausible range of margins. Whatever the conversion turns out to be, the calibrated elements sit about a factor 13 above the turn. Everywhere on the range of Figure 2, and everywhere the calibrated elements sit, the mixed derivative is therefore negative in both classes, and what separates them is the price ratio reported in Section 3.2.

A.6 Preferences

The reversal and the $\epsilon^{-1/2}$ exponent are stated under log utility, which gives the exact value-function separation of Section 2.1.

A.6.1 Epstein–Zin at unit elasticity

Everything rests on the consumption rule, so check it first. At unit elasticity of substitution the aggregator is

$$f(c, J) = \rho(1-\gamma)\,J\left[\log c - \tfrac{1}{1-\gamma}\log\big((1-\gamma)J\big)\right], \tag{32}$$

and with $J = \frac{K^{1-\gamma}}{1-\gamma} e^{(1-\gamma)w(M,X)}$ the first-order condition $\partial_c f = J_K$ returns $c^* = \rho K$, as under log and independent of γ and of the climate state. The reduction to a single equation in the carbon–climate block therefore goes through.

That equation gains one term. Substituting $c^* = \rho K$ and dividing by $K^{1-\gamma} e^{(1-\gamma)w}$ leaves

$$\rho w = \rho \log \rho + (A-\rho) + \max_{a\in[0,1]} \big\{ -C(a) + w_M (1-a) E_0 \big\} + f(X;M)\, w_X + \tfrac{\sigma^2}{2} w_{XX} + \underbrace{\tfrac{\sigma^2}{2}(1-\gamma)\, w_X^2}_{\text{risk sensitivity}}, \tag{33}$$

the log problem plus a quadratic term that vanishes at $\gamma = 1$: a risk-averse planner penalises climate-state variance, a risk-tolerant one rewards it. The price is $\mathrm{SCC} = -K\, w_M$.

One substitution removes that term. Writing $\beta = 1-\gamma$ and $\Psi = e^{\beta w}$ collapses diffusion and risk sensitivity into a single second derivative, and factoring out the no-tipping value, $\Psi = \Psi_0 \chi$ with $\rho w_0 = S_0$, cancels the source terms and leaves

$$\frac{\sigma^2}{2} \chi_{XX} + f\, \chi_X + (1-a) E_0\, \chi_M - \rho\, \chi \log \chi = 0, \qquad \chi(x_-) = e^{-\beta \mathcal{L}}, \quad \chi \to 1. \tag{34}$$

Two consequences. The third term is the transport term of Section 3.1.2 and drops under the same condition that licenses the split there; and since $w = w_0 + \beta^{-1} \log \chi$, the price separates into a Ramsey and a threshold component here as it does under log.

The exponent survives. Measuring the state in layer units, $x = \sigma^{2/3} \eta$ and $\mu = \sigma^{4/3} \nu$, removes σ from (34) except through $\tilde{\rho}$,

$$\tfrac{1}{2} \chi'' + (\nu - \eta^2) \chi' - \tilde{\rho}\, \chi \log \chi = 0, \tag{35}$$

so the moving boundary, the square-root margin and the layer width all follow from drift against diffusion, none of which involves γ. Risk aversion reaches the amplitude and never the rate at which the margin closes, which is what fixes the power, $\mathrm{SCC}_{\text{threshold}} \sim A_\gamma\, \epsilon^{-1/2}$.

What changes is the level. Equation (35) is nonlinear, where the log problem was linear and left its solution known only up to a constant; here the boundary value $e^{-\beta\mathcal{L}}$ pins that constant, so the loss enters the amplitude as well as multiplying it,

$$A_\gamma = \mathcal{L} \cdot \mathsf{f}(\rho T) \cdot \frac{\Phi_\gamma(\tilde{\rho}, \beta \mathcal{L})}{\sigma^{2/3} \sqrt{M^*}},$$

and the price stops being proportional to the loss away from $\gamma = 1$, its elasticity exceeding one below unit risk aversion and falling short above, by about ten percent per unit of $|\beta\mathcal{L}|$ at the calibrated $\tilde{\rho}$. Proportionality is a property of logarithmic utility rather than of the model, and a usable approximation where we calibrate.

The reversal needs the amplitude positive and increasing. Positivity is a theorem: writing $V = -\log \chi / \beta \mathcal{L}$ puts (35) in a form the maximum principle signs, so $\Phi_\gamma > 0$ for either sign of β. Monotonicity is numerical, over four decades bracketing the calibrated values and for $|\beta\mathcal{L}|$ up to eight, with no turning point, and for the fold alone. What the paper claims under these preferences is the exponent, and nothing downstream rests on the reversal holding away from $\gamma = 1$.

A.6.2 Power utility

Here the separation fails, and it fails structurally. Optimal consumption is $c^* = K\,G^{-1/\gamma}$, so it responds to the climate state, and the reduced equation for G carries a nonlinearity, $\gamma\,G^{(\gamma-1)/\gamma}$, that no factorisation removes. The no-tipping solution no longer factors out, so the price does not separate into a Ramsey component and a threshold component that can be computed apart and added. The term that blocks it carries no small parameter, so there is no proximity at which the split becomes usable.

That failure also decides what kind of evidence is available. Epstein–Zin at unit elasticity keeps $c^* = \rho K$ and a reduced problem from which σ scales out exactly, so there the exponent is derived. Here it has to be measured.

Table 5: Log-log price slopes from the full value-function problem, solved separately for each γ at $\sigma = 0.20$, $\rho = 0.04$, $A = 0.02$, $L = 0.50$ and $x_{\max} = 6.0$, fitted by least squares over $\epsilon \in [5 \times 10^{-6}, 10^{-3}]$. All four lie within 2.2% of -0.500.

γ	Slope	$\lvert\text{err}\rvert$	Interpretation
0.5	-0.512	2.4%	EIS $= 2$ (high substitution)
1.0	-0.497	0.6%	Log utility (exact Prop. Lemma)
2.0	-0.492	1.6%	Standard CRRA
4.0	-0.490	2.0%	High risk aversion
Theory	-0.500	exact	Perturbation argument above

The slope comes back within 2.2% of $-\frac{1}{2}$ at every γ tested, on a window three orders of magnitude below the calibrated proximities. The exponent therefore outlives the derivation that produced it. The reversal holds throughout as well: sweeping σ over $[0.10, 0.50]$ at fixed proximity, the price falls strictly with the noise at every $\gamma \in \{0.5, \ldots, 6\}$.

A.7 Numerical methods and validation

The body and Appendix A.3 establish the divergence exponent, the noise reversal and the two-class structure analytically. What remains is computation: the values of the amplitude and the numerical confirmation of the two-class slope.

A.7.1 The solver

Two boundary-value problems are solved, one linear and one not, and the same fourth-order collocation scheme on an adaptive mesh handles both. The linear one is the discounted first-passage equation (4) with $\varphi(x_-) = 1$ and $\varphi \to 0$; the price is read off $|\partial\varphi/\partial\epsilon|$ at the stable state and the amplitude as the small-μ limit of $(1 - \varphi(x^*))/\sqrt{\mu}$ rescaled by $\sigma^{2/3}$.

The nonlinear one is the power-utility analogue of Section A.6.2, solved in a normalised unknown $\tilde{u}$ that leaves every ratio the paper reads off it unchanged,

$$\frac{\sigma^2}{2}\tilde{u}'' + (\mu - x^2)\,\tilde{u}' + b\big(\tilde{u}^{\,(\gamma-1)/\gamma} - \tilde{u}\big) = 0, \qquad \tilde{u} \in (0,1), \tag{36}$$

with $\tilde{u}(-\sqrt{\mu};\mu) = 1 - L$ and $\tilde{u}(x_{\max};\mu) = 1$, and $b = \rho - A(1-\gamma)$.

Two checks anchor the scheme, and both test a prediction the analysis makes rather than internal consistency. The first is that the amplitude depends on σ and ρ only through $\tilde{\rho}$, which is hypothesis (H2): across a grid of (σ, ρ) at fixed $\tilde{\rho}$ the extracted $\Phi_{\text{fold}}/\sigma^{2/3}$ is stable to within 0.04%, collapsing onto a single curve. The second is that the log-log slope of the price against the remaining budget converges to -0.500 for the divergent geometries, which it does, with the leading formula in relative error below 5% for $\epsilon < 0.05$ and the residual scaling as $\sqrt{\epsilon}$ as (30) predicts.

A.7.2 Values of the amplitude

Appendix A.3 fixes the sign, the monotonicity and the slope at the origin; the solver supplies the values in between. Table 6 reports them at parameters relevant to the climate case, which sits at $\tilde{\rho} \geq 0.14$, low enough that the small-$\tilde{\rho}$ column of Table 4 is the one to read, though not so low that the amplitude is exactly linear there.

Table 6: The amplitude $\Phi_{\text{fold}}(\tilde{\rho})$, computed from the first-passage problem (4).

$\tilde{\rho} = \rho/\sigma^{2/3}$	0.00	0.05	0.10	0.20	0.50	1.00	2.00
Φ_{fold}	0.000	0.098	0.190	0.355	0.754	1.232	1.888

A.7.3 The two-class slope

Figure 4 confirms the divergence. For the fold and for the pitchfork we solve (4) on a fine grid of ϵ, read off the price, and take the limiting slope of its logarithm against $\log \epsilon$. The diagnostic is that limiting slope, and $-1/2$ is the signature of the divergence: both geometries lock onto it, the pitchfork more slowly because its higher-order corrections are larger.

A.8 Other bifurcations

A.8.1 Hopf bifurcation

Everywhere else the sign change of Section 2.5 is established by scaling: far from the threshold the crossing value becomes exponentially small and the sign turns. The Hopf lets it be seen instead. Where the noise dominates, the cubic drift is negligible and the crossing value depends on the oscillation's amplitude alone, so the problem becomes one-dimensional in the radius and its decaying solution is standard, $K_0(\alpha r)$. Noise enters through one length only, $1/\alpha = \sigma/\sqrt{2\rho}$, the distance over which the crossing value falls away, so the elasticity of the price to the noise is

$$E_\sigma = \alpha r^* \frac{K_0(\alpha r^*)}{K_1(\alpha r^*)} - \alpha r_c \frac{K_1(\alpha r_c)}{K_0(\alpha r_c)}, \tag{37}$$

with r^* the stable amplitude and r_c the critical one below which the oscillation dies out. As the threshold is approached, $r^* \to r_c$ and the two terms become a ratio and its own

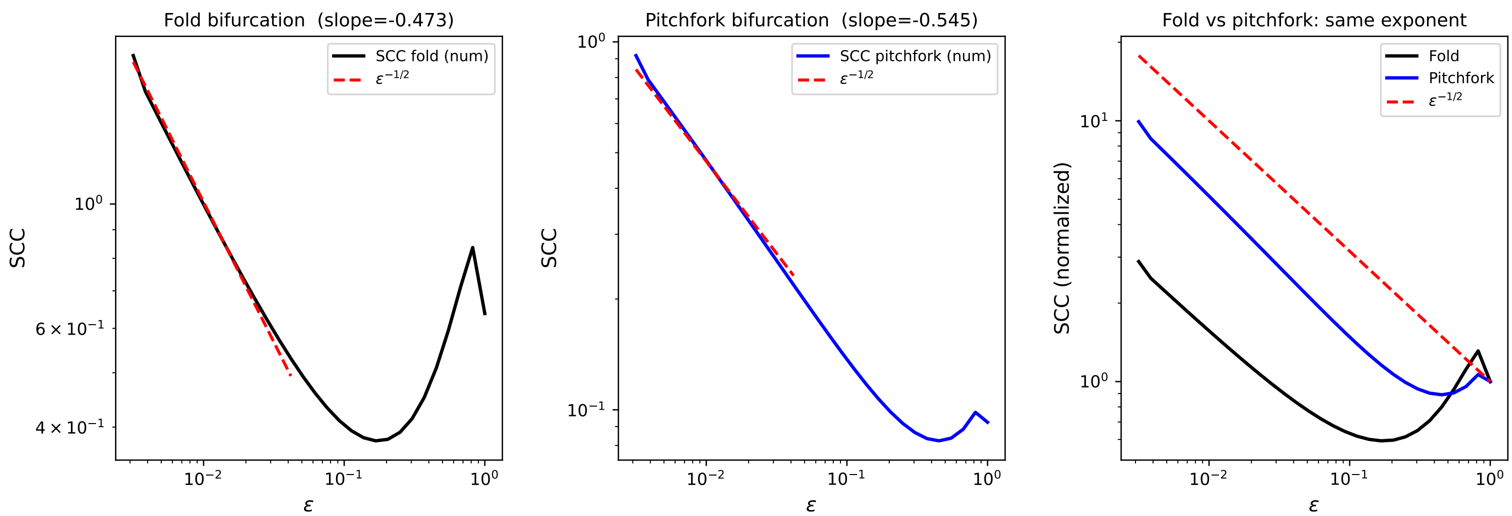

Figure 4: Numerical confirmation of the divergent class, from direct solution of (4) at $\sigma = 1$ and $\rho = 0.05$. Left and centre: the log-log plot of $\mathrm{SCC}_{\mathrm{threshold}}(\epsilon)$ against ϵ for the fold and for the pitchfork, each against an $\epsilon^{-1/2}$ reference. Right: the two prices normalised at the right-hand edge and superposed, which is where the rate of approach can be read, the pitchfork reaching the common slope more slowly.

reciprocal. That ratio is below one at every argument, so the difference is negative and the reversal holds. Far from the threshold the first term dominates and the elasticity is positive. The window is therefore not an artefact of the expansion, and its two ends are elementary.

The Hopf also places itself in the bounded class. With $\dot{r} = r(\epsilon - r^2)$ the oscillation settles at $r^* = \sqrt{\epsilon}$, the threshold is reached at $\epsilon_c = r_c^2$, and the margin $d = \sqrt{\epsilon} - \sqrt{\epsilon_c}$ closes linearly, so $p = 1$. Noise and impatience do not collapse into $\tilde{\rho}$ here, so neither conclusion is a corollary: the reversal comes from (37) and the class from the margin law above.

A.8.2 Pitchfork bifurcation

The pitchfork canonical form is $dX = (\mu X - X^3)\, dt + \sigma\, dW$, with stable equilibria $x^* = \pm\sqrt{\mu}$ and an unstable boundary at the origin, so the margin $d(\mu) = \sqrt{\mu}$ collapses as a square root: Class A, with a price diverging as $\epsilon^{-1/2}$. Its amplitude is not a second object but Φ_k at $k = 3$, so Section A.3.5 supplies its positivity, monotonicity and concavity along with the fold's. Solving the crossing problem on the right basin, the numerical log-log slope converges to -0.500 as the threshold is approached, more slowly than the fold because its higher-order corrections are larger (Figure 5, left).

A.8.3 Transcritical bifurcation

The transcritical canonical form is $dX = (\mu X - X^2)\, dt + \sigma\, dW$, with stable equilibrium $x^* = \mu$ and boundary at the origin, so the margin $d(\mu) = \mu$ collapses linearly: Class B.

Nothing further needs computing, because this drift is the fold's translated: the crossing value is the same function of the same arguments, and the transcritical therefore carries the

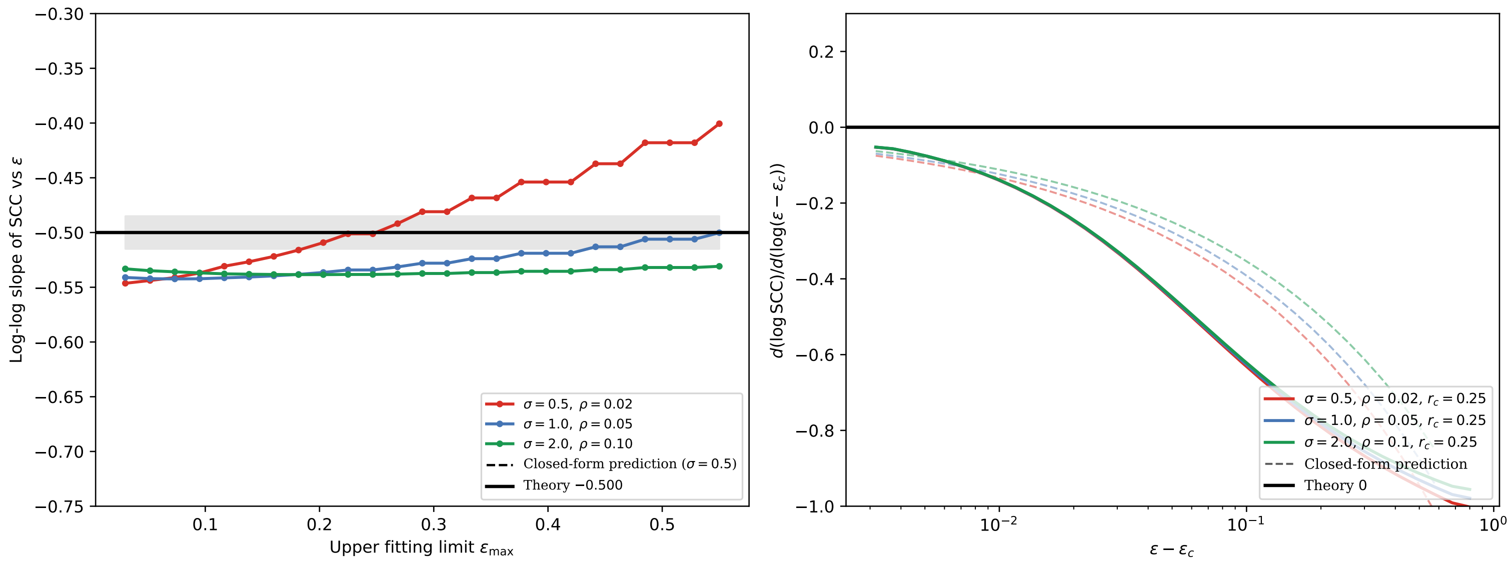

Figure 5: Numerical slope diagnostics from first-passage BVP solutions. Left: the pitchfork log-log slope of SCC versus ϵ, against the fitting upper limit $\epsilon_{\max}$, for three calibrations; the slopes converge to -0.500 as $\epsilon_{\max} \to 0$ (Class A). Right: the Hopf slope versus $(\epsilon - \epsilon_c)$ for $r_c = 0.25$, converging to 0 as $\epsilon \to \epsilon_c$ (Class B).

fold's steepness $\Phi_{\text{fold}}(\tilde{\rho})/\sigma^{2/3}$ exactly rather than approximately. What differs is how the stable equilibrium moves with the budget, linearly here against $1/(2\sqrt{\mu})$ for the fold, so there is no divergence and

$$\mathrm{SCC}_{\text{tc}}(\epsilon) \;\to\; \frac{\mathcal{L}\,\Phi_{\text{fold}}(\tilde{\rho})}{\sigma^{2/3}}, \qquad \epsilon \to 0. \tag{38}$$

A.9 Two coupled elements

The body treats tipping elements as independent. Coupling does not change the class: for a pair of coupled fold elements the $\epsilon^{-1/2}$ exponent survives, the partner acting only through a shift of the effective proximity and a small amplification. This holds under two conditions on the partner, which we state rather than derive.

Consider two coupled fold elements,

$$dX_1 = \left(\mu_1 - X_1^2 + \gamma_1 X_2\right) dt + \sigma_1\, dW_1, \tag{39}$$

$$dX_2 = \left(\mu_2 - X_2^2 + \gamma_2 X_1\right) dt + \sigma_2\, dW_2, \tag{40}$$

with independent W_1, W_2. Element 1 now tips at a boundary that moves with X_2, and its stable state solves $X_1^* = \sqrt{\mu_1 + \gamma_1 X_2^*}$ against $X_2^* = \sqrt{\mu_2 + \gamma_2 X_1^*}$. The budget that governs element 1 is therefore the square of that state, and the natural variable is

$$\epsilon_{1,\text{eff}} \;=\; \frac{X_1^{*2}}{M_1^*} \;=\; \frac{\mu_1 + \gamma_1 X_2^*}{M_1^*}, \tag{41}$$

equal to ϵ_1 when the coupling vanishes and larger under positive coupling, the partner holding element 1 further from its fold.

The two conditions are on element 2. It must stay below its own threshold, so that X_2^* exists and is stable. And it must be quiet and fast relative to element 1: its fluctuations, once scaled by the coupling, small against element 1's margin, and its own relaxation quick enough to track its fixed point while element 1 approaches its. Where either fails the partner is not a shifted budget but a second slow variable, and the one-dimensional reduction of Section 2.5 no longer applies.

Granting them, the partner reaches element 1 only through $\mu_{1,\text{eff}} = X_1^{*2}$, its noise absorbed into that shift at leading order. The single-element analysis applies with ϵ_1 replaced by $\epsilon_{1,\text{eff}}$, and differentiating through the coupled fixed point supplies one extra factor, the response of the effective budget to the real one, $\Lambda = d\mu_{1,\text{eff}}/d\mu_1$:

$$\text{SCC}_1^{\text{coupled}} = \frac{\mathcal{L}}{M_1^*} \cdot \frac{\Phi_{\text{fold}}(\tilde{\rho}_1)}{\sigma_1^{2/3}\sqrt{M_1^*}} \cdot \left(\epsilon_{1,\text{eff}}\right)^{-1/2} \cdot \Lambda, \qquad \Lambda = \frac{4X_1^* X_2^*}{4X_1^* X_2^* - \gamma_1\gamma_2}. \tag{42}$$

The exponent is preserved exactly: read against the effective proximity, the log-log slope is $-1/2$. The factor Λ collects the feedback that returns through the partner, and equals one in a one-way cascade, where $\gamma_1\gamma_2 = 0$.

Numerically the shift dominates and the amplification does not. At the coupling strengths reported for AMOC–Amazon interaction, $|\gamma| \lesssim 0.15$, the amplification is below three percent; solving the coupled first-passage problem, the log-log slope against the effective proximity is -0.504, against -0.327 for the naive ϵ_1, so reading the price against the right variable restores the exact exponent (Figure 6). Those coupling strengths fix an order of magnitude rather than price that pair, since both elements sit outside the window of Section 2.5; what the exercise establishes is the structure of the price under coupling.

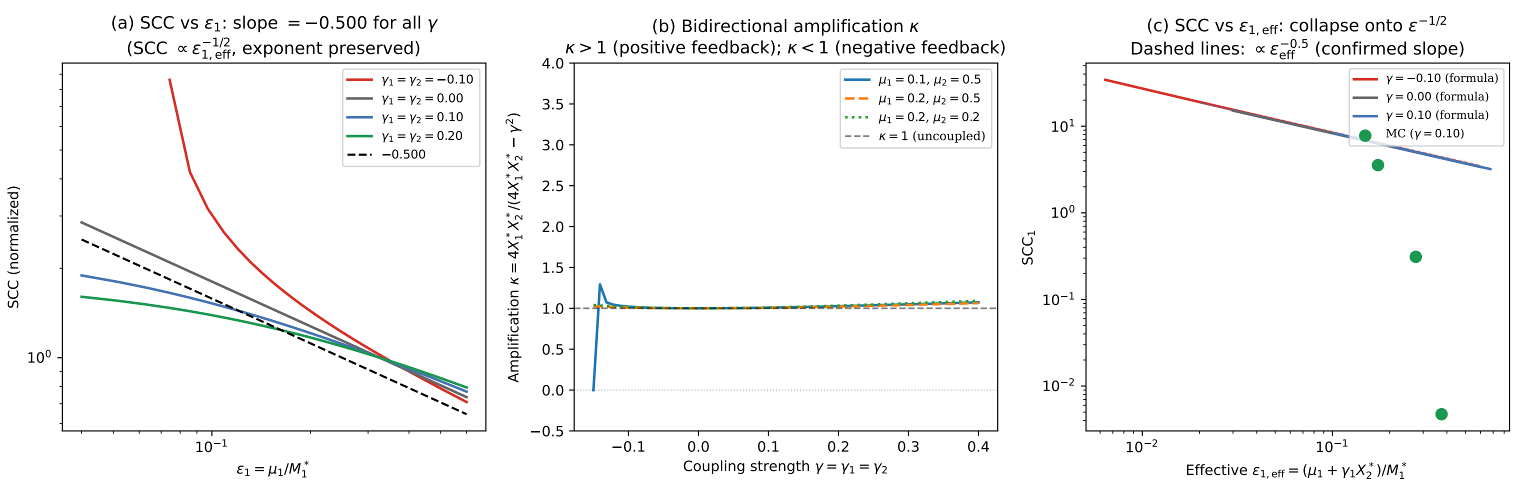

Figure 6: Coupled fold elements. Left: SCC_1 against the naive proximity ϵ_1 for several coupling strengths, giving a family of parallel curves. Right: the same prices against the effective proximity (41), collapsing onto one curve.