University of California, Berkeley
Department of Agricultural &
Resource Economics

CUDARE Working Papers

Tipping Points and Ambiguity in the Economics of Climate Change

Derek M. Lemoine
and Christian Traeger

Tipping Points and Ambiguity
in the Economics of Climate Change[*]

Derek M. Lemoine[†] & Christian Traeger[§]

[†] Department of Economics, University of Arizona
McClelland Hall 401, 1130 E Helen St, Tucson, AZ, 85721-0108, USA
dlemoine@email.arizona.edu

[§] Department of Agricultural & Resource Economics, University of California, Berkeley
207 Giannini Hall #3310, Berkeley, CA 94720-3310, USA
traeger@berkeley.edu

We model optimal policy when the probability of a tipping point, the welfare change due to a tipping point, and knowledge about a tipping point's trigger all depend on the policy path. Analytic results demonstrate how optimal policy depends on the ability to affect both the probability of a tipping point and also welfare in a post-threshold world. Simulations with a numerical climate-economy model show that possible tipping points in the climate system increase the optimal near-term carbon tax by up to 45% in base case specifications. The resulting policy paths lower peak warming by up to 0.5°C compared to a model without possible tipping points. Different types of tipping points have qualitatively different effects on policy, demonstrating the importance of explicitly modeling tipping points' effects on system dynamics. Aversion to ambiguity in the threshold's distribution can amplify or dampen the effect of tipping points on optimal policy, but in our numerical model, ambiguity aversion increases the optimal carbon tax.

JEL Codes: Q54, D90, D81

Keywords: tipping point, threshold, regime shift, ambiguity, climate, uncertainty, integrated assessment, dynamic programming, social cost of carbon, carbon tax

1   Introduction

Potentially irreversible shifts in system dynamics play a key role in important economic applications such as the macroeconomic consequences of bankruptcies, institutional changes resulting from mass uprisings, and management of environmental systems subject to abrupt degradation. How these shifts affect welfare depends on the policy decisions made before and after tipping occurs. Crucially for policy, however, the thresholds that trigger such tipping points are generally uncertain: the policymaker often trades immediate payoffs against the chance of tipping the dynamic system into a new regime. Further, while a policymaker can learn about a threshold's location over time, the probabilities governing tipping points are rarely known with confidence. The policymaker therefore often also trades more confidently known payoffs against less confidently known payoffs. We capture both types of trade-offs in a dynamic model of optimal policy. We analyze the determinants of policy when choices affect both the probability of a tipping point and its welfare impact. By combining our analysis of uncertain tipping points with a model of ambiguity, we also analyze how optimal policy changes when a decision-maker is more averse to tipping point uncertainty than to more confidently known risks.

Our numerical application explores how tipping points in the climate system affect optimal carbon taxes over time. Scientists have grown increasingly concerned about the possibility of abrupt changes caused by crossing temperature thresholds (Alley et al., 2003; Overpeck and Cole, 2006; Lenton et al., 2008). While these concerns have shaped the policy debate, economic models for estimating the social cost of carbon have only allowed for smooth and reversible changes in the climate system (Greenstone et al., 2011). In our base case specifications, tipping points raise the current optimal carbon tax by around 40%, with ambiguity aversion raising the tax further. We explore qualitative policy differences arising from different tipping points, and we demonstrate the importance of explicitly modeling how dynamics shift and also how policy affects the probability of a shift.

Our model endogenizes the timing and probability of crossing a threshold. Crossing a threshold causes an irreversible change in system dynamics, which we call a tipping point.[1] We also endogenize both the welfare change induced by crossing a threshold and learning

---

[1]A different notion of tipping point sometimes used in the economics literature refers to shifts between equilibria due to small changes in parameters. These changes can occur due to preferences in residential sorting models (Schelling, 1971), network externalities in technology adoption models (Katz and Shapiro, 1994), increasing returns in agglomeration models (Ellison and Fudenberg, 2003), and human capital accumulation in growth models (Azariadis and Drazen, 1990). The non-convex control literature directly models the possibility of tipping a system into a new type of equilibrium (e.g., Skiba, 1978; Brock and Starrett, 2003; Mäler et al., 2003; Wagener, 2003). In these deterministic models, it is always known whether a given policy path will or will not tip the system, so tipping along the optimal path depends entirely on the initial conditions. Skiba points divide the space of initial conditions into regions with and without optimal tipping.

about the threshold location. Previous work has not jointly endogenized these tipping point characteristics. The macroeconomic literature studies the potential for monetary policy to stabilize the economy when nominal interest rate rules might shift discontinuously in the future (Davig and Leeper, 2007), and the real options literature studies optimal investment when demand dynamics might shift discontinuously in the future (Guo et al., 2005). Because these regime shifts are controlled by exogenously fixed transition probabilities, the decision-maker can only change the welfare impact of tipping (self-insure) but not change the likelihood of tipping (self-protect). The resources literature allows policy to affect the probability of an irreversible "climate catastrophe," but the flow of disutility due to the catastrophe is specified exogenously (Clarke and Reed, 1994; Tsur and Zemel, 1996; Gjerde et al., 1999; Nævdal, 2006; Nævdal and Oppenheimer, 2007). In contrast, the welfare changes in our model derive from the explicit changes in system dynamics and depend on how optimal policy adjusts to tipping points.

More closely related work in the resource literature develops models that shift system dynamics upon crossing a threshold (Heal, 1984; Brozović and Schlenker, 2011; Polasky et al., 2011; de Zeeuw and Zemel, 2011). Our analytic model is more general than these stylized models, adds learning and ambiguity, and provides new insights into the equilibrium conditions that drive optimal policies in the face of tipping points. In particular, as in Tsur and Zemel (1996), learning enables our decision-maker to realize that already-explored regions of the state space are free of thresholds. The most closely related numerical work is by Keller et al. (2004), who models a threshold that alters ocean circulation. In contrast to their approach, we use a recursive dynamic programming framework that enables us to include endogenous learning and annual stochasticity. Further, we directly model the effect of a tipping point on climate dynamics rather than approximating its effects by shifting the damage function.

The probability of incurring a tipping point is less understood than are the distributions governing many stochastic processes. For instance, in our climate application, the probability of crossing a temperature threshold is less confidently known than is interannual temperature variability. We employ a recent model of ambiguity attitude based on Klibanoff et al. (2005, 2009) and Traeger (2010) to express the decision-maker's relative lack of confidence in threshold probabilities as opposed to the probabilities governing more standard risks.[2] We demonstrate that ambiguity aversion generally has an "ambiguous" effect on optimal policies and analyze the determinants of its impact. In our numeric climate change application, we find that ambiguity aversion slightly increases the optimal carbon tax. Previous

---

[2]Keynes (1921), Knight (1921), and Ellsberg (1961) each suggested distinguishing confidence in different types of uncertainty.

work on ambiguity in the climate change context either uses abstract two-period models or evaluates exogenous policy paths (Lange and Treich, 2008; Millner et al., 2010). We are not aware of any work analyzing ambiguity in the context of tipping points or analyzing smooth ambiguity aversion in a dynamic, optimizing integrated assessment model.

Sections 2 and 3 introduce the general model and analytically describe how anticipating tipping points affects optimal policy. Sections 4 and 5 present our numerical integrated assessment model of climate change and show the expected optimal carbon tax and optimal temperature path. Section 6 discusses implications of learning and system inertia for optimal policy paths. We conclude in Section 7. The online appendix provides the complete model description, additional results, and derivations.

2   Modeling tipping points

Our tipping points are irreversible shifts in system dynamics that occur upon crossing a threshold in the state space. The policymaker does not know the precise location of the threshold. The probability of a tipping point occurring (i.e., the hazard rate) is endogenous. It depends on the evolution of the state variables, which in turn depend on policy choices as well as on the stochastics governing system dynamics. The policymaker learns that regions in the state space she has already visited are free of tipping points.[3] Crossing the threshold shifts the world from the "pre-threshold" regime to a "post-threshold" regime with permanently altered system dynamics. Optimal pre- and post-threshold policies together determine the welfare loss triggered by the tipping point.

The policymaker solves an infinite-horizon dynamic optimization problem. Optimal policy at time t depends on the vector S_t of state variables. We denote the value of the optimal policy program by $V_\psi(S_t)$.[4] The parameter ψ indicates whether V is the value function for the pre-threshold regime ($\psi = 0$) or for the post-threshold regime ($\psi = 1$). In general, the threshold is an unknown function of the state variables. In the case of climate change, the threshold is the temperature level. Once the threshold is crossed, system dynamics change irreversibly. Returning state variables to earlier values does not restore the original dynamics. In our climate application, the new dynamics include melted ice sheets, large methane releases, or disrupted forest ecosystems; lowering temperature would not undo any of these

---

[3]The more of the relevant state space that is already explored without crossing the threshold, the more likely that the threshold is in the remaining unexplored state space. Some region of the relevant state space might not be explored even under policies optimized without considering tipping point possibilities. The probability mass on the permanently unexplored region can be interpreted as the chance that there is no tipping point.

[4]In our numerical application, each value function $V_\psi(\cdot)$ will be non-stationary. We absorb this non-stationarity in the the state vector by making time t a component of S_t.

changes over policy-relevant timescales. Similarly, macroeconomic changes can permanently alter expectations and institutions. Optimal policy in the pre-threshold regime must consider its effect on both the pre- and post-threshold value functions, but once the state variables cross the threshold, optimal policy depends only on post-threshold dynamics. Therefore, we solve the model recursively, starting with the post-threshold problem and then substituting the solution into the pre-threshold problem.

In the post-threshold world, we obtain our value (and policy) functions from solving the following Bellman equation:

$$V_{\psi=1}(S_t) = \max_{x_t} \left\{ u(x_t, S_t) + \beta_t \int V_{\psi=1}(S_{t+1}) \, d\mathbb{P} \right\} \tag{1}$$

$$\text{s.t. } S_{t+1} = g_{\psi=1}(x_t, \epsilon_t, S_t)$$

$$x_t \in \Gamma(S_t).$$

Here, x_t is the vector of time t control variables, $u(\cdot)$ is the utility derived from time t consumption, and β_t is the discount factor.[5] Constraints on the controls are captured by the set $\Gamma(S_t)$. The transition function $g_{\psi=1}(\cdot)$ characterizes post-threshold dynamics. At time t, the next period's state vector is S_{t+1}. It depends on the vector ϵ_t of independently and identically distributed stochastic shocks whose distribution is characterized by the probability measure $\mathbb{P}$. The decision-maker maximizes the sum of immediate utility and discounted expected future welfare. The value function $V_{\psi=1}$ is defined as the fixed point of equation (1) and determines welfare after crossing the threshold. The welfare change from crossing the threshold therefore depends on the state variables at the time of crossing and, thus, on the policy path chosen prior to crossing.

Prior to crossing a threshold, the value of the optimal policy program is given by the pre-threshold value function:

$$V_{\psi=0}(S_t) = \max_{x_t} \left\{ u(x_t, S_t) + \beta_t \int \left[[1 - h(S_t, S_{t+1})] \, V_{\psi=0}(S_{t+1}) \right.\right.$$

$$\left.\left. + h(S_t, S_{t+1}) \, V_{\psi=1}(S_{t+1}) \right] d\mathbb{P} \right\} \tag{2}$$

$$\text{s.t. } S_{t+1} = g_{\psi=0}(x_t, \epsilon_t, S_t)$$

$$x_t \in \Gamma(S_t).$$

---

[5]The dependence of the discount factor β_t on time can arise as a consequence of reformulating utility and normalizing consumption and state variables. In our climate change application, we use effective labor units for consumption and capital and transform the population-weighted utility function of per capita consumption into the form stated above (Crost and Traeger, 2010).

The pre-threshold value function captures the possibility of crossing the threshold. The endogenous hazard $h(S_t, S_{t+1})$ determines the risk of crossing the threshold between the current and the next period. This probability generally depends on the current state variables and on how they change from one period to the next.[6] With probability $1 - h$, the system dynamics stay unaltered and $V_{\psi=0}$ characterizes future welfare. With probability h, the system tips and $V_{\psi=1}$ determines future welfare from period $t+1$ on. Because of the stochasticity in the equations of motion, we take expectations over the next period's value functions and over the hazard rate (via the integral). Once we have solved equation (1) for $V_{\psi=1}$, we find $V_{\psi=0}$ as the fixed point of equation (2).

Finally, we generalize the welfare evaluation in equation (2) to recognize that we often know little about the distribution governing the threshold location. By assumption, we have no records of having recently crossed a threshold, and abrupt changes in system dynamics usually do not lend themselves to accurate forecasting. Tipping point possibilities therefore exhibit a deeper uncertainty in the sense that the probabilities governing tipping points are less confidently known than are well-measured probability distributions. The smooth ambiguity model captures this distinction between well-known probabilities, here our annually observed stochasticity captured by ϵ_t, and more subjective and less confidently known uncertainty, here our hazard rate h. An ambiguity-averse decision-maker is more averse to the less confidently known tipping point uncertainty than to the risk posed by the random shock. To capture ambiguity aversion, we generalize the pre-threshold welfare evaluation:

$$V_{\psi=0}(S_t) = \max_{x_t} \left\{ u(x_t, S_t) + \beta \int f_{amb}^{-1} \left[[1 - h(S_t, S_{t+1})] \ f_{amb}[V_{\psi=0}(S_{t+1})] \right. \right.$$
$$\left. \left. + h(S_t, S_{t+1}) \ f_{amb}[V_{\psi=1}(S_{t+1})] \right] d\,\mathbb{P} \right\} \qquad (3)$$
$$\text{s.t. } S_{t+1} = g_{\psi=0}(x_t, \epsilon_t, S_t)$$
$$x_t \in \Gamma(S_t) .$$

The concave function f_{amb} captures smooth ambiguity aversion (Klibanoff et al., 2005, 2009), or intertemporal risk aversion to subjective uncertainty (Traeger, 2010). When f_{amb} is linear, equation (3) reduces to the ambiguity-neutral form in equation (2). Behavioral evidence supports a concave f_{amb} as a descriptive model of many decision-makers (Camerer and Weber, 1992). Further, when the von Neumann-Morgenstern axioms are extended to recognize that a decision-maker can have differing confidence in different distributions, concave f_{amb} is

---

[6]In general, the state space contains informational variables that tell the decision-maker which part of the state space has already been explored. In our climate application, the decision-maker keeps track of the greatest historic temperature.

consistent with normatively attractive preferences (Traeger, 2010).

3 The effects of tipping points on optimal policy

We now identify the channels by which tipping points affect optimal policy. The possible existence of a tipping point introduces two new terms into the marginal welfare impact of changing a control. For ease of exposition, we analyze the case where a single state variable determines the chance of crossing the threshold. The right-hand side of equation (3) characterizes welfare for an optimal choice of the controls (with optimality denoted by *). We evaluate the marginal welfare impact of varying a generic entry e_t of the control vector in the neighborhood of the optimum. In our climate change application, the temperature state variable determines the hazard, and the welfare impact of varying emissions determines the optimal carbon tax. Suppressing all arguments independent of e_t, the value of the optimal policy program is:

$$u(e_t^*) + \beta_t \int \underbrace{f_{amb}^{-1}\left[\left[1 - h(S_{t+1}(e_t^*))\right]f_{amb}\left[V_{\psi=0}(S_{t+1}(e_t^*))\right] + h(S_{t+1}(e_t^*))f_{amb}\left[V_{\psi=1}(S_{t+1}(e_t^*))\right]\right]}_{V_{eff}(e_t^*)} d\mathbb{P}.$$

The integrand $V_{eff}(e_t^*)$ expresses the value of future periods' optimal policy program in utility units. It is the ambiguity-averse mean of pre- and post-threshold value. The total value aggregates V_{eff} over temperature risk and combines it with the utility from current consumption. Varying e_t gives us the following trade-off characterizing optimal policies:

$$
\begin{aligned}
\frac{\partial u(e_t^*)}{\partial e_t} = -\beta \int \Bigg\{ &\left[1 - h(S_{t+1}(e_t^*))\right]\frac{f_{amb}'\left[V_{\psi=0}(S_{t+1}(e_t^*))\right]}{f_{amb}'\left[V_{eff}(e_t^*)\right]}\frac{\partial V_{\psi=0}(S_{t+1}(e_t^*))}{\partial S_{t+1}}\frac{\partial S_{t+1}(e_t^*)}{\partial e_t} \\
&+ h(S_{t+1}(e_t^*))\frac{f_{amb}'\left[V_{\psi=1}(S_{t+1}(e_t^*))\right]}{f_{amb}'\left[V_{eff}(e_t^*)\right]}\frac{\partial V_{\psi=1}(S_{t+1}(e_t^*))}{\partial S_{t+1}}\frac{\partial S_{t+1}(e_t^*)}{\partial e_t} \\
&- \underbrace{\frac{\partial h(S_{t+1}(e_t^*))}{\partial S_{t+1}}}_{(i)}\underbrace{\frac{\partial S_{t+1}(e_t^*)}{\partial e_t}}_{(ii)}\underbrace{\frac{f_{amb}[V_{\psi=0}(S_{t+1}(e_t^*))] - f_{amb}[V_{\psi=1}(S_{t+1}(e_t^*))]}{f_{amb}'\left[V_{eff}(e_t^*)\right]}}_{(iii)} \Bigg\} d\mathbb{P} \,,
\end{aligned}
$$

(4)

where primes ($'$) indicate derivatives.[7] We interpret this equation for the case where an increase in e_t raises current utility but decreases expected future welfare. For instance,

---

[7] For a multidimensional state space, $\partial V/\partial S_{t+1}$ and $\partial h/\partial S_{t+1}$ denote gradients and $\partial S_{t+1}/\partial e_t$ denotes the vector of state changes caused by the marginal change in e_t. These derivatives are taken with respect to the pre-threshold dynamics $g_{\psi=0}$.

additional carbon dioxide emissions increase current utility but decrease future welfare by generating higher carbon stocks and temperatures; additional borrowing increases current consumption but also increases future debt; and a political elite's appropriation of resources and repression increase social discontent and hamper growth. We assume for now that f_{amb} is the identity function, which implies an ambiguity-neutral decision-maker.

The left-hand side of equation (4) characterizes the (immediate) benefits from increasing the policy variable. At the optimum, these benefits must balance the expected future costs. The costs are represented by the right-hand side of equation (4) and are subject to uncertainty (captured by the integration). The integrand in the first line represents the impact of policy on time $t+1$ welfare under the pre-threshold regime (i.e., on the pre-threshold continuation value). This impact is composed of the control's impact on the state vector and the effect of the altered state vector on pre-threshold value $V_{\psi=0}$, and it is weighted by the probability of staying in the pre-threshold regime $(1-h)$. In a world without tipping points (where the hazard rate h is zero), the first line characterizes the full trade-off between current and future welfare.

The possibility of imminent tipping points introduces the second and third lines into the optimal policy trade-off: altering e_t now also changes time $t+1$ welfare in the post-threshold world (second line) and changes the probability of entering the post-threshold world (third line). The first two lines together give the expected marginal welfare effect of increasing e_t in situations where the immediate hazard rate h is exogenous. These lines adjust a model without tipping points to account for the different marginal effect of the control e_t on pre- and post-threshold welfare. We therefore call this adjustment the *differential welfare impact (DWI)*:

$$DWI^{neutral} \equiv h \left(\frac{\partial V_{\psi=0}}{\partial S_{t+1}} - \frac{\partial V_{\psi=1}}{\partial S_{t+1}} \right) \frac{\partial S_{t+1}(e_t^*)}{\partial e_t} \, , \tag{5}$$

where we maintain the assumption of ambiguity neutrality. The DWI is proportional to the hazard rate and to the difference in the marginal impact of the control on the pre- and post-threshold value functions. If increasing the control decreases welfare relatively more in the post-threshold regime, then the differential welfare impact makes raising the control more costly.

The third line in equation (4) only arises when the tipping point's probability is endogenous. In this case, a change in the control e_t affects the hazard rate. The optimal policy now has to account for this marginal change in the hazard rate in response to a change in the control. We call this contribution the *marginal hazard effect (MHE)*. The MHE is composed of the response of the hazard rate to a change in the state vector (term i), the response of the

state vector to a change in the control (term ii), and the total welfare change from switching regimes (term iii). For the tipping points in our climate application, increasing emissions raises the hazard rate and the welfare difference $[V_{\psi=0} - V_{\psi=1}]$ is always positive. Therefore, the MHE increases the cost of emissions relative to a case without tipping points. In general, a change in the current control could also increase hazard rates at future times. Possible future tipping points are included in the pre-threshold continuation value $V_{\psi=0}$. If the current control alters the probability of crossing a threshold in the future, then $\partial V_{\psi=0}/\partial S_{t+1}$ will include this effect.

In summary, anticipating possible tipping points adjusts the first-order conditions governing optimal policy for the differential impact of the control on pre- and post-threshold welfare (DWI) and for the control's effect on the immediate hazard rate (MHE). In our climate change application, the effects of DWI and MHE together increase the optimal carbon tax. If we instead consider additional borrowing by a highly indebted state, additional debt increases the risk of altering investors' beliefs about solvency and so tipping into a credit regime with high interest rates on sovereign debt (MHE). This change in dynamics can lead to default. In October of 2011, Eurozone leaders managed the post-tipping dynamics by negotiating "haircuts" on Greek debt. However, this type of post-threshold management can result in a DWI effect that reduces a country's incentive to avoid the tipping point. If other highly indebted countries in the monetary union count on a similar post-threshold "haircut", their cost of taking up additional debt is reduced because, in case of tipping, they only have to pay back a fraction of the debt. Here the DWI acts as a benefit, reducing the cost of the tipping point for each individual nation.

In another example, a political elite that appropriates its country's resources has to account for the increased risk of an uprising that puts an end to its rule (MHE). We consider two possible post-threshold scenarios. In the first, a successful uprising puts a clear end to power and wealth. Then the DWI merely adjusts the elite's calculus by eliminating payoffs in the event of losing power. It increases the cost of tipping and, in general, reduces its likelihood. This scenario is equivalent to a model with constant post-threshold utility and endogenous tipping risk. In the second, a successful uprising puts an end to the elite's power but does not affect its previously accumulated wealth. If money and power are substitutes, the value of a marginal unit of wealth would be larger in the post-threshold regime. The DWI now works in the opposite direction from the first scenario, increasing the likelihood of a tipping point.

We now consider the consequences of ambiguity aversion. An ambiguity-averse decision-maker is additionally averse to the types of poorly understood uncertainty that characterize tipping points. A strictly concave function f_{amb} captures her additional aversion to tipping

point uncertainty. Our extended definition of DWI now collects all the changes in the first two lines of equation (4) with respect to the continuation value:

$$
DWI^{total} = DWI^{neutral} + \overbrace{(1-h)\left(\frac{f'_{amb}(V_{\psi=0})}{f'_{amb}(V_{eff})}-1\right)\left(-\frac{\partial V_{\psi=0}}{\partial e_t}\right)}^{DWI^{ambig}}
$$
$$
+ \underbrace{h\left(\frac{f'_{amb}(V_{\psi=1})}{f'_{amb}(V_{eff})}-1\right)\left(-\frac{\partial V_{\psi=1}}{\partial e_t}\right)}_{DWI^{ambig}}. \tag{6}
$$

These changes capture the effect of an exogenous hazard and its interaction with ambiguity aversion. In our interpretation, we assume that increasing the control (e.g., emitting an additional unit) reduces future welfare more in the post-threshold regime than in the pre-threshold regime ($DWI^{neutral} > 0$).[8]   For an ambiguity-averse decision-maker who faces a welfare-decreasing tipping point, the concavity of f_{amb} implies $f'(V_{\psi=0}) < f'(V_{eff}) < f'(V_{\psi=1})$. Therefore, the first term's contribution to DWI^{ambig} in equation (6) is negative, while the second term's contribution is positive. For a small hazard rate and a second-order expansion of f_{amb} (see appendix), we find

$$
DWI^{ambig} \approx \left.\frac{-f''_{amb}}{f'_{amb}}\right|_{V_{eff}} [V_{\psi=0}-V_{\psi=1}] \overbrace{h\left(\frac{\partial V_{\psi=0}}{\partial e_t}-\frac{\partial V_{\psi=1}}{\partial e_t}\right)}^{DWI^{neutral}}
$$
$$
- \frac{1}{2}\left.\frac{-f''_{amb}}{f'_{amb}}\right|_{V_{eff}}\left.\frac{-f''_{amb}}{f'_{amb}}\right|_{V_{\psi=0}} [V_{\psi=0}-V_{\psi=1}]^2 \; h \; \left(-\frac{\partial V_{\psi=0}}{\partial e_t}\right). \tag{7}
$$

The ambiguity effect captured in the first line increases the ambiguity-neutral DWI in proportion to the measure of absolute ambiguity aversion $\left.-f''_{amb}/f'_{amb}\right|_{V_{eff}}$. The measure of absolute ambiguity aversion is zero for an ambiguity-neutral decision-maker and is positive for an ambiguity-averse decision-maker. The effect captured in the second line relates closely to how ambiguity aversion adjusts the pre-threshold value function. When $DWI^{neutral} > 0$, the first line increases the marginal cost of the control while the second line decreases it. In our climate change application, the second line's contribution is often the larger one, making ambiguity aversion decrease the optimal carbon tax.

Ambiguity aversion also affects the marginal hazard effect (MHE) via term iii in the third

---

[8]In our climate application, the current control interacts with future threshold crossings even if a threshold is not crossed in the current period. Damages increase faster with emissions in the post-threshold world, which raises DWI. However, at the same time pre-threshold emissions can trigger future tipping via the delay in the temperature equation (see appendix), which decreases the DWI. Our simulations suggest that these conflicting effects do produce $DWI^{neutral} < 0$ in some cases.

line of equation (4):

$$\frac{f_{amb}[V_{\psi=0}] - f_{amb}[V_{\psi=1}]}{f'_{amb}[V_{eff}]} = [V_{\psi=0} - V_{\psi=1}] \underbrace{\frac{f_{amb}[V_{\psi=0}] - f_{amb}[V_{\psi=1}]}{[V_{\psi=0} - V_{\psi=1}] f'_{amb}[V_{eff}]}}_{\text{ambiguity multiplier}} . \tag{8}$$

The first term in brackets is the ambiguity-neutral contribution. The fraction characterizes the ambiguity contribution. It is unity for an ambiguity-neutral decision-maker and when $f'_{amb}[V_{eff}] = \frac{f_{amb}[V_{\psi=1}] - f_{amb}[V_{\psi=0}]}{[V_{\psi=1} - V_{\psi=0}]}$. Given the concavity of f_{amb}, this condition has to be satisfied for some V_{eff} and, thus, for some hazard h. If the hazard is lower than this critical value, then $f'_{amb}[V_{eff}]$ is also lower and ambiguity aversion amplifies the MHE. This insight relates closely to the common finding that ambiguity aversion increases the effective weight of low-probability catastrophic events (pessimism bias). If the hazard is greater than the critical value, ambiguity aversion reduces MHE. The intuition is familiar from the standard risk setting: a risk-averse agent is not always willing to pay more for a risk reduction than is a risk-neutral agent.[9] Here, the MHE captures willingness to pay for a marginal reduction in the tipping point hazard. In our climate change application, the annual hazard rate is small, leading ambiguity aversion to always increase the MHE and so the optimal carbon tax.

In general, the effect of increasing ambiguity aversion is less easily determined than that of a high or low hazard rate. Increasing ambiguity aversion for a given hazard rate has two effects. First, increasing ambiguity aversion reduces the mean value V_{eff}, decreasing the ambiguity multiplier. Second, increasing ambiguity aversion changes the derivative f'_{amb} at a given value (relative to the secant). For the case of small hazard rates as in our climate change application, a first-order expansion in h pins down the effect of increasing ambiguity aversion. Employing once more a second-order expansion of f_{amb} (see appendix), we obtain the following approximation of the ambiguity multiplier:

$$\frac{f_{amb}[V_{\psi=0}] - f_{amb}[V_{\psi=1}]}{[V_{\psi=0} - V_{\psi=1}] f'_{amb}[V_{eff}]} \approx$$

$$1 + \underbrace{\left.\frac{-f''_{amb}}{f'_{amb}}\right|_{V_{eff}} (V_{\psi=0} - V_{\psi=1})}_{\text{order zero}} \underbrace{\left(\frac{1}{2} - h\left[1 + \frac{1}{2} \left.\frac{-f''_{amb}}{f'_{amb}}\right|_{V_{\psi=0}} [V_{\psi=0} - V_{\psi=1}]\right]\right)}_{\text{first-order correction}} . \tag{9}$$

For a sufficiently small hazard, the ambiguity multiplier increases proportionally to the mea-

---

[9]The more risk-averse (ambiguity-averse) the decision-maker, the more he values wealth in bad states relative to good states. Giving up wealth for a hazard reduction makes an ambiguity-averse decision-maker even worse off if the bad outcome happens. If a hazard is large enough and costly enough to reduce, he prefers carrying wealth into the post-threshold world over spending it on reducing the hazard.

sure of absolute ambiguity aversion. As the hazard becomes larger, the first-order correction reduces the ambiguity effect. The hazard rate in our climate change application is sufficiently small that the first-order correction is negligible. On the whole, we find that ambiguity aversion alters both channels by which tipping points affect optimal policy. Because our climate change application has a small annual hazard, ambiguity aversion primarily increases the ambiguity-neutral MHE in proportion to the measure of absolute ambiguity aversion.

4   A climate-economy model with tipping points, learning, and ambiguity

We now consider the effect of climate tipping points on the optimal carbon tax. The optimal carbon tax equals the social cost of carbon when evaluated along the optimal policy path.

We reformulate the benchmark Dynamic Integrated model of Climate and the Economy (DICE) from Nordhaus (2008) as an infinite-horizon dynamic programming problem with a tipping point in the climate system, optimal learning about the threshold that triggers a tipping point, and a generalized welfare evaluation.[10] DICE is a Ramsey-Cass-Koopmans growth model that has an aggregate world economy interacting with a climate module (Figure 1). Gross economic output (or potential GDP) is determined by an endogenous capital stock, an exogenously growing labor force, and exogenously improving production technology. Gross output produces carbon dioxide (CO_2) emissions. Non-abated CO_2 emissions accumulate in the atmosphere and ultimately translate into global warming, which causes damage proportional to world output. Cumulative temperature change affects the total output available for allocation by the policymaker. The control variables are abatement and consumption, and residual output not allocated to these two options becomes capital investment. The state variables are capital per effective unit of labor, the stock of CO_2 in the atmosphere, the change in global mean surface temperature since 1900, and, to keep track of exogenously evolving variables, time.

A tipping point irreversibly changes the climate system from its conventional representation in DICE to a new regime with altered dynamics. The tipping point occurs upon crossing some unknown temperature threshold. Emission decisions determine the future CO_2 stock, thereby affecting future temperatures and the probability that a tipping point occurs. The decision-maker anticipates how he would choose emissions and consumption in the post-threshold world. The timing, probability, and welfare consequences of a regime switch are

---

[10]The appendix provides the model equations. The standard DICE model is a nonlinear programming problem with constant system dynamics. For previous work using recursive versions of DICE, see Kelly and Kolstad (1999), Leach (2007), and Crost and Traeger (2010).
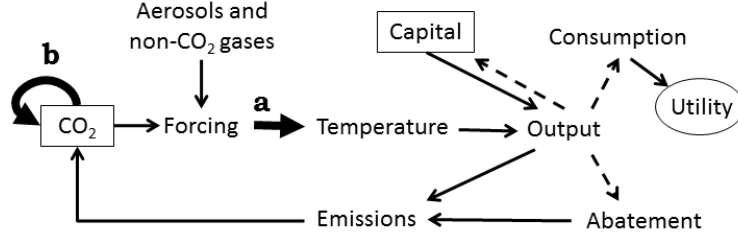
Figure 1: A simplified schematic of the relation between the economy and the climate. Boxes indicate stock variables, and dashed arrows indicate the decision variables of consumption, investment, and abatement. The climate feedback tipping point makes temperature more sensitive to forcing (a), and the carbon sink tipping point increases the persistence of CO_2 (b).

endogenous because they depend on the policies chosen before and after the threshold occurs.

To model tipping points, we specialize the recursive structure from Section 2 to DICE. We have one dynamic programming problem for the post-threshold world and another for the pre-threshold world. The pre-threshold world has standard DICE dynamics along with the tipping possibility and temperature shocks calibrated to the historical record. A tipping point produces the post-threshold world by irreversibly changing the standard dynamics. We first solve the post-threshold problem and then use its solution in the pre-threshold problem. We numerically solve each dynamic programming problem for the unknown value function using function iteration. Employing a projection method, we approximate the value functions by Chebychev polynomials and use collocation at the Chebychev nodes in the four-dimensional state space (Miranda and Fackler, 2002).

We evaluate two tipping points of prominent concern in the climate science literature.[11] In every model run, the policymaker faces a single tipping point and knows in advance what its effects would be. The first tipping point increases the climate feedbacks that amplify global warming (arrow a in Figure 1), and the second increases the atmospheric lifetime of CO_2 (arrow b in Figure 1). The first tipping point therefore increases the effect of emissions on temperature, and the second increases the time during which emissions affect the climate. The climate science literature has compiled a number of pathways by which tipping points could abruptly change the strength of feedbacks that determine surface temperature. As one example, warming could mobilize large methane stores locked in permafrost and in ice lattices (clathrates) in the shallow ocean (Hall and Behl, 2006; Archer, 2007; Schaefer et al., 2011). If warming mobilizes these methane stocks, they would cause further warming that could

---

[11]Each modeled tipping point is an extreme case: climate dynamics change severely, abruptly, and irreversibly. Science does not offer clear guidance on the best way to model a given tipping point, so we translate two common tipping stories into DICE's reduced climate system in order to gain intuition about policy implications.

mobilize additional stocks. As another example, if land ice sheets begin to retreat on decadal timescales, the resulting loss of reflective ice could double the long-term warming predicted by models that hold land ice sheets fixed (Hansen et al., 2008). Temperature dynamics in DICE depend on a parameter known as climate sensitivity, which is the equilibrium warming from doubling CO_2. The value of 3°C used in DICE is inferred from climate models that hold land ice sheets and most methane stocks constant. We represent a climate feedback tipping point as increasing climate sensitivity to 4°C, 5°C, or 6°C.

The second tipping point reflects the possibility that carbon sinks weaken beyond the predictions of coupled climate-carbon cycle models. Warming-induced changes in oceans (Le Quéré et al., 2007), soil carbon dynamics (Eglin et al., 2010), and standing biomass (Huntingford et al., 2008) could affect the uptake of CO_2 from the atmosphere. We represent these weakened sinks by decreasing the transfer of CO_2 out of the atmosphere by 25%, 50%, or 75%. The reader may think of this tipping point as reducing the "decay rate" of atmospheric CO_2. A given emission path produces a higher CO_2 concentration once this threshold is crossed. Further, if the threshold triggers a strong form of this tipping point, then the flow of carbon from land and ocean sinks back into the atmosphere can temporarily outweigh the flow of carbon out of the atmosphere. The resulting negative decay rate represents the system coming to a new equilibrium with more CO_2 in the atmosphere.

The system passes from the pre-threshold regime ($\psi_t = 0$) into the post-threshold regime ($\psi_{t+1} = 1$) when cumulative temperature change T_{t+1} crosses an unknown threshold $\tilde{T}$. Every temperature between the maximum temperature previously reached and an upper bound $\bar{T}$ has an equal chance of being the threshold, meaning $\tilde{T}$ is uniformly distributed between the historic maximum and $\bar{T}$.[12] In our base case model runs, we use $\bar{T} = 4.27$°C so that the year 2005 expected value for the threshold is 2.5°C.[13] Sensitivity analyses vary $\bar{T}$ between 3°C and 9°C, implying year 2005 expected values of about 1.9°C to 4.9°C. The probability of crossing the threshold between periods t and $t + 1$ conditional on not having crossed the

---

[12]The optimal policy path in the absence of tipping points reaches a maximum temperature of 3.33°C. Our model with $\bar{T} > 3.33$ is therefore equivalent to one with the uniform distribution's upper bound at 3.33°C and probability $(\bar{T} - 3.33)/(\bar{T} - T_t)$ that there is no threshold.

[13]Using E_{2005} $\tilde{T} = 2.5$°C is consistent with the political 2°C limits for avoiding dangerous anthropogenic interference. Further, in Smith et al. (2009), 2.5°C is in the upper end of the temperature region that produces significant risk of large-scale discontinuities and is just below the temperatures that produce severe risk.

threshold by time t is:[14]

$$h(T_t, T_{t+1}) = \max\left\{0, \frac{\min\{T_{t+1}, \bar{T}\} - T_t}{\bar{T} - T_t}\right\} \ . \tag{10}$$

This expression is the hazard of crossing the tipping point. As the world reaches higher temperatures without reaching a threshold, the decision-maker learns that the threshold is above the current temperature and updates his beliefs by moving probability density from the newly safe region to the remaining unexplored temperatures. Therefore, as the world safely reaches higher temperatures, each unit of temperature increase creates a greater hazard than it did at lower temperatures. The state variable T_t that controls the threshold crossing is a climate variable whose equation of motion is determined by CO_2 concentrations and does not reflect annual stochastic fluctuations.

As noted in discussing equation (3), the function f_{amb} captures smooth ambiguity aversion (Klibanoff et al., 2005, 2009), or intertemporal risk aversion under subjective uncertainty (Traeger, 2010). We use the term "subjective uncertainty" to describe uncertain outcomes when there is less information available for determining probabilities. This deficiency in probabilistic knowledge applies to climate tipping points, which are less understood than other climate phenomena (Alley et al., 2003; Lenton et al., 2008; Ramanathan and Feng, 2008; Kriegler et al., 2009; Smith et al., 2009). Sticking to isoelastic preferences, we adopt the power function $f_{amb}(V) = ((1 - \eta)V)^{\frac{1-\gamma}{1-\eta}}$. Here γ is a measure of Arrow-Pratt relative risk aversion with respect to subjective or poorly understood uncertainty and η is the constant Arrow-Pratt measure of relative risk aversion used in the standard DICE utility function. The subjective uncertainty contrasts with the standard risk posed by the historically-grounded temperature stochasticity and governed by $\mathbb{P}$ (see appendix). If Arrow-Pratt risk aversion with respect to subjective risk γ coincides with standard risk aversion η, the function f_{amb} is linear and drops out. In that case, the policymaker is ambiguity-neutral and the welfare evaluation is as in DICE. However, when $\gamma > \eta$, the function f_{amb} measures the policymaker's additional aversion to ambiguity (or subjectivity of belief) governing tipping points as opposed to annual temperature stochasticity.

---

[14]In DICE-2007, the CO_2 stock increases monotonically until the model reaches a sufficiently high level of abatement. From this point on, the "decay" of CO_2 outweighs the flow of emissions, making the CO_2 stock decrease monotonically. Temperature in DICE follows the same pattern. When temperature is increasing, the probability of crossing the threshold is proportional to the difference between the next period's temperature and the current temperature. When temperature is decreasing, the probability of crossing the threshold is 0. As long as temperature is a quasiconcave function of time, we do not need an additional state variable to keep track of the highest historic temperature.

5   The optimal carbon tax when facing possible tipping points

We compare several sets of model runs to assess how the optimal carbon tax responds to the type of tipping point considered, to the strength of a tipping point, to prior beliefs about the temperature threshold location, and to aversion to tipping point ambiguity. All of our graphs present results conditional on not having crossed the threshold: we want to understand how optimal policy changes in the face of a potential tipping point. The depicted paths draw the multiplicative temperature shock at its expected value in each period. Each graph compares the baseline scenario without tipping point awareness to runs with tipping points of various strengths. The appendix contains additional results.

Figure 2 gives the effect of tipping points on the optimal carbon tax (social cost of carbon), the optimal CO_2 concentration path, and the optimal temperature path. The figure assumes ambiguity neutrality and the base case prior over the threshold location. The year 2015 optimal carbon tax is near \$10/tCO_2 in the absence of tipping points, the strongest version of the feedback tipping point increases it to \$13.5/tCO_2, and the strongest version of the carbon sink tipping point increases it to \$14/tCO_2. While tipping point possibilities have only a modest effect on near-term abatement, they can nonetheless have a large effect on cumulative abatement because they increase the optimal tax by proportionally more later in the century.[15] The optimal path without possible tipping points produces a peak temperature (CO_2 concentration) of 3.3°C (637 ppm), reached in the year 2187 (2163). The optimal tax path in the presence of the weak climate feedback tipping point reduces this peak temperature to 3.0°C (592 ppm), while the higher taxes justified by the strong climate feedback tipping point further reduce peak temperature to 2.8°C (560 ppm). The tax path in the presence of the weak carbon sink tipping point reduces peak temperature only to 3.2°C (617 ppm), while the possibility of the strong carbon sink tipping point reduces peak temperature to 3.0°C (588 ppm). By reducing peak temperature and CO_2, the decision-maker reduces the cumulative probability of crossing the temperature threshold. The stronger the anticipated tipping point, the more output the decision-maker devotes to reducing this probability.

Because we explicitly model the effects of tipping points on system dynamics, different types of tipping points can have qualitatively different effects on optimal policy. In comparison to the carbon sink tipping points, the climate feedback tipping points affect the optimal carbon tax relatively less in the near-term but relatively more later in the century.

---

[15]The climate feedback tipping points have their greatest proportional effect on the optimal carbon tax shortly after 2100, while the proportional effect of the carbon sink tipping points peaks shortly after 2050. The appendix plots how abatement and other variables respond to each tipping point possibility.
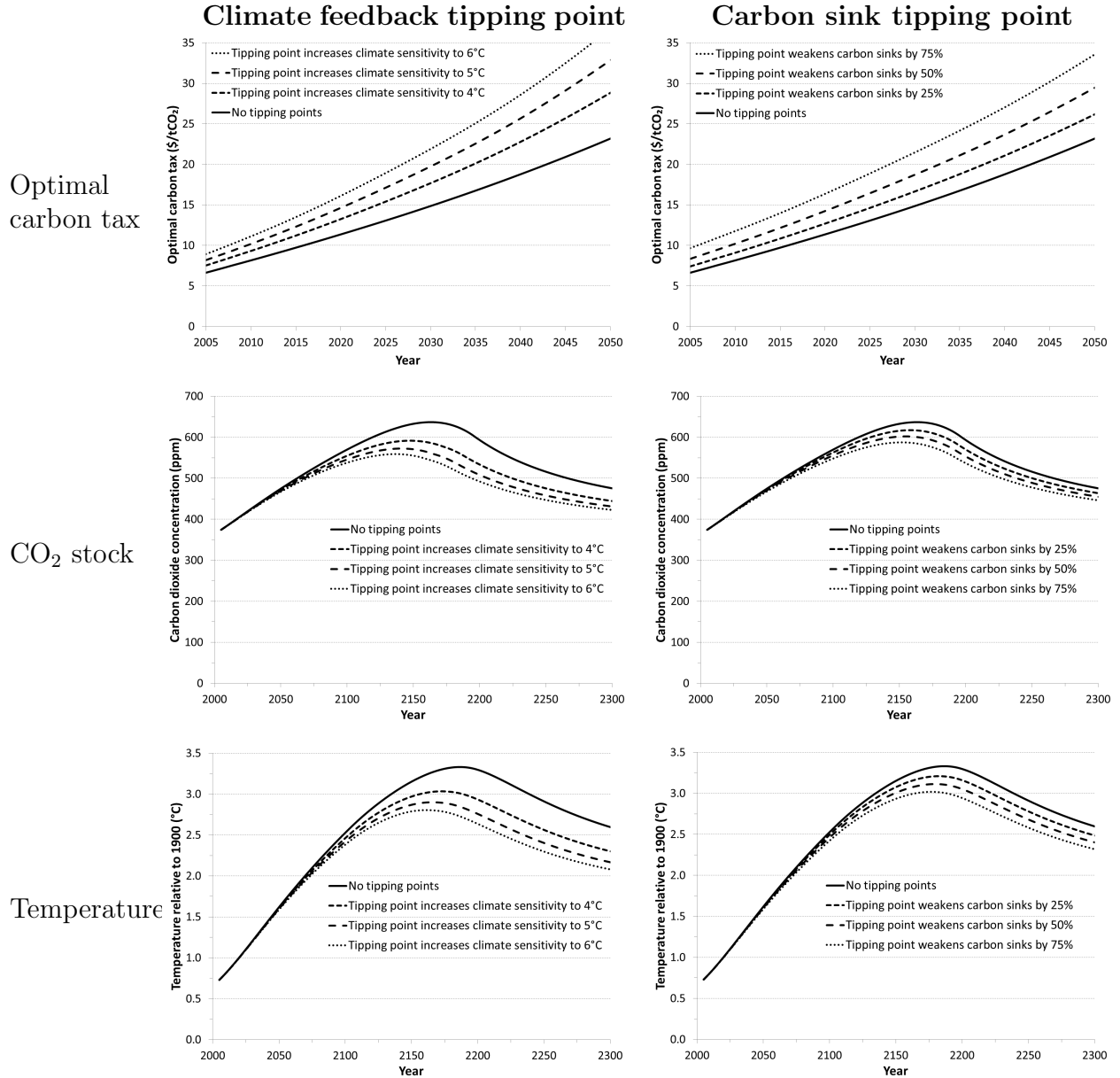
Figure 2: Time paths for the optimal carbon tax (current value), the CO_2 stock, and temperature under each type of tipping point possibility using expected draws. We simulate a path that happens to never cross a threshold in order to see how the modeled policymaker adjusts to the possibility over time. Results are for an ambiguity-neutral policymaker with $\bar{T} = 4.27°C$.

The main reason for this difference is that the MHE increases much more sharply in the CO_2 concentration when facing a climate feedback tipping point. Once the tipping point's additional climate feedbacks are activated, a high CO_2 concentration immediately implies a higher climate forcing and a faster warming process. The policymaker responds by sharply increasing abatement. If the additional feedbacks are activated early in the current century, optimal policy produces a much lower CO_2 path. However, when the feedbacks are activated later in the century, they act upon a higher CO_2 concentration. This higher CO_2 level produces greater temperature change and also greater loss of welfare. In contrast, once carbon sinks have degraded, CO_2 persists longer in the atmosphere and can temporarily increase even in the absence of further emissions. Crossing the carbon sink tipping point produces higher long-run concentrations whether the threshold is crossed when CO_2 is low or high. Welfare loss and MHE therefore depend relatively less on the CO_2 concentration at the point of crossing than in the case of a feedback tipping point. If a threshold is crossed in the near-term, carbon sink tipping points reduce welfare more sharply than do feedback tipping points, but the welfare cost of a feedback tipping point grows faster as CO_2 increases. These differences imply that a possible carbon sink tipping point has a relatively stronger impact on policy early in the current century while a possible feedback tipping point has a relatively stronger impact on policy later in the century.

Recognizing the present inability of climate science to provide a probability distribution for the temperature threshold, we now consider the implications of more and less diffuse priors for the threshold location and of aversion to the ambiguity in the threshold's distribution. Figure 3 plots the year 2015 optimal carbon tax and the peak temperature for values of $\bar{T}$ between 3°C and 9°C, with all calculations still being for optimal policy paths under ambiguity neutrality and conditional on not having crossed the threshold. As the upper bound $\bar{T}$ increases, optimal policy converges asymptotically to the scenario without a tipping point. A more diffuse prior on the threshold location reduces the importance of the tipping point contributions in equation (4) by reducing both the hazard rate and its derivative. Lowering $\bar{T}$ from its base case value has a stronger effect on optimal policy than does raising $\bar{T}$. For low values of $\bar{T}$, the hazard rate is a steeper function of emissions (raising MHE) and realized temperatures can approach regions with a high hazard rate (raising DWI). When $\bar{T} = 3$°C, the optimal carbon tax in 2015 rises as high as \$16.5/tCO$_2$ for the strong tipping points, with temperature peaking just above 2.5°C.

Finally, we turn to the effect of ambiguity aversion on optimal policy. Figure 4 varies γ from ambiguity-neutral ($\gamma = 2 = \eta$) to extremely ambiguity-averse ($\gamma = 100 > 2 = \eta$). We find that optimal policy is not highly sensitive to the policymaker's level of ambiguity aversion. The near-term carbon tax varies by less than \$2/tCO$_2$ across the modeled range
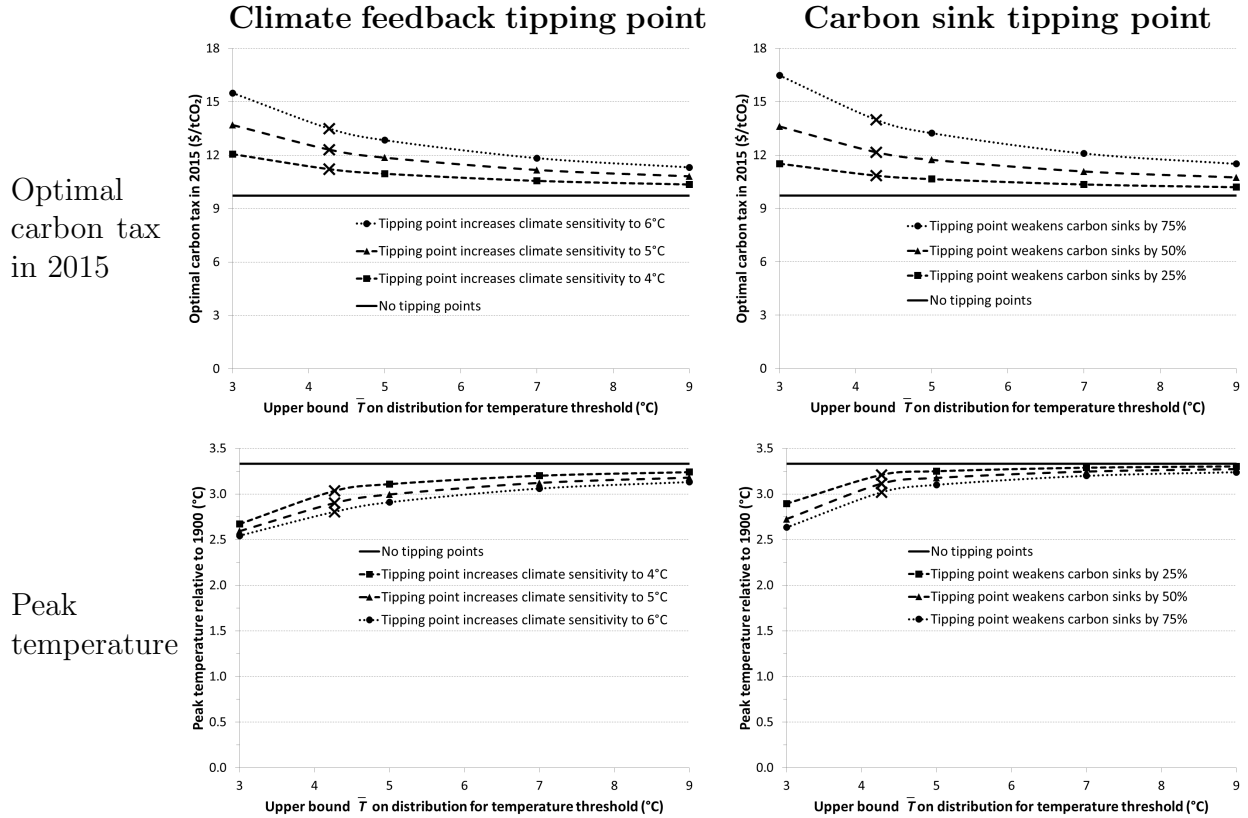
Figure 3: The optimal carbon tax in 2015 and the peak temperature reached for each upper bound $\bar{T}$ for the temperature threshold's distribution. The plotted simulations assume expected draws of the temperature shock, assume that the tipping point never occurs, and model an ambiguity-neutral policymaker. Points marked by X use the base case distribution.
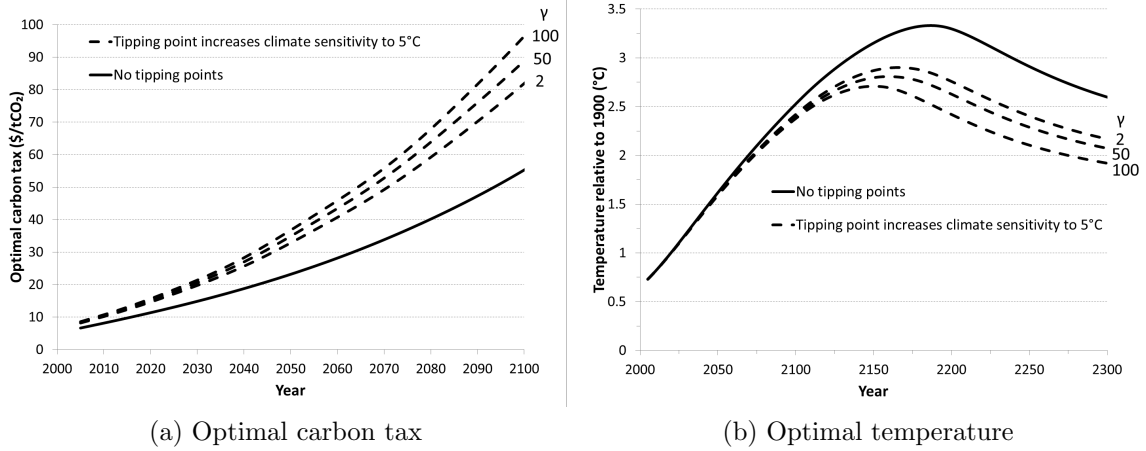
(a) Optimal carbon tax

(b) Optimal temperature

Figure 4: The optimal carbon tax and the optimal temperature for different degrees of aversion (γ) to threshold uncertainty. Crossing the threshold triggers the mid-strength climate feedback tipping point. An ambiguity-averse policymaker has $\gamma > 2$. The plotted simulations assume expected draws of the temperature shock and also assume that the tipping point never occurs.

of γ, the peak CO_2 concentration varies by less than 20 ppm, and peak temperature varies by less than 0.2°C. Ambiguity aversion decreases DWI and raises MHE. However, because of the small annual hazard, its effect on MHE is more significant. Equation (9) shows that this amplification is proportional to the total welfare loss from crossing the threshold and to the measure of absolute ambiguity aversion (near 3 for $\gamma = 100$). Ambiguity aversion therefore increases the MHE by more for stronger tipping points that have more severe welfare implications, and the effect of ambiguity aversion on the optimal tax grows as the welfare loss from crossing a tipping point grows over time. For the middle climate feedback tipping point, the extreme form of ambiguity aversion with $\gamma = 100$ increases the optimal carbon tax by 6% in 2015 and by 12% in 2050. For the middle carbon sink tipping point, the extreme form of ambiguity aversion also increases the optimal carbon tax by 6% in 2015, but increases it by only 7% in 2050.[16] Tipping point evaluation has to acknowledge that threshold distributions are, by necessity, guesstimates rather than objective probabilities. In our application, ambiguity aversion always increases the social cost of carbon. The effect on policy of varying ambiguity aversion across our scenarios is slightly smaller than the effect of varying the strength of a tipping point.

---

[16]Note from Figure 4 that the effect of greater ambiguity aversion becomes more pronounced later in the century for the climate feedback tipping point. With the carbon sink tipping point, the approximate constancy of the ambiguity effect over time means that greater ambiguity aversion has about the same effect late in the century as in the beginning of the century.

6   Policy implications of system inertia in the presence of learning

This section explains how inertia in the climate system changes the profile of the optimal temperature and policy trajectories. Our policymaker is a Bayesian learner when it comes to the location of an irreversible threshold. We first describe qualitative implications of learning in general optimally controlled systems with irreversible tipping points. We then highlight the importance of explicitly capturing inertia when undertaking quantitative analysis of optimal climate policy. In particular, when information about threshold locations depends on the evolution of state variables, the form of the optimal policy trajectory is sensitive to how precisely immediate policy controls the next period's state variables. While most deterministic integrated assessments of climate change capture inertia, many models with uncertainty or learning do not.[17] The basic insight underlying this section is that, in the absence of inertia, it is eventually optimal to maintain temperature for a long time at the highest level learned to be safe. However, with inertia in the system, temperature only peaks once the CO_2 concentration is already falling. The policy that would keep temperature constant must change significantly from period to period, making it inefficient to keep temperature constant for an extended time.

Optimal policy sets emissions so that marginal abatement cost equals marginal damage in each period. Because marginal abatement cost falls exogenously over time in DICE, abatement increases over time. Tipping point possibilities increase marginal damage through the MHE, but as long as temperature has been monotonically increasing, learning makes the MHE depend on whether the next period's temperature is above the current period's temperature. The marginal damage curve therefore has a discontinuity at the emission level $\hat{e}(M, T, t)$ that keeps temperature constant in a period.[18] For emissions above $\hat{e}$ (point a in Figure 5a), marginal damage includes the effect of emissions on tipping point possibilities. At $\hat{e}$, however, marginal damage depends on whether it is assessed relative to the next unit of emissions or relative to the last unit of emissions: the damages caused by the *next* abated unit are strictly higher than the cost of abating that unit, but abatement cost is in turn strictly higher than the damages caused by the *last* abated unit (point b). The policymaker therefore holds emissions fixed at $\hat{e}$. As abatement cost falls, emissions move to a part of the

---

[17]For instance, of the work in the resources literature that studies catastrophes or shifts in dynamics, Gjerde et al. (1999), Keller et al. (2004), and Nævdal and Oppenheimer (2007) model inertia. The majority do not model inertia (e.g., Heal, 1984; Clarke and Reed, 1994; Tsur and Zemel, 1996; Nævdal, 2006; Brozović and Schlenker, 2011; Polasky et al., 2011; de Zeeuw and Zemel, 2011).

[18]The emission level that keeps temperature constant depends on current temperature and CO_2 because of the climate system's inertia (discussed below) and depends on time because of exogenous variables that affect forcing and non-industrial emissions (see appendix).

marginal damage curve that does not include threshold possibilities (points c and d) because temperature stays in the region that has been shown to be safe.

The possibility of learning about which temperatures are safe from thresholds makes optimal policy more sensitive to how temperature evolves over time. When temperature is governed by a delay equation as in DICE (see appendix), the emission level that holds temperature constant can change significantly from period to period. Optimal policy would therefore have to jump dramatically in order to hold temperature constant. However, if temperature were determined wholly by the current period's CO_2 stock, then the policymaker would more precisely control temperature in each period. Eliminating inertia in temperature would make it optimal to stay right at the boundary of the safe region by holding temperature constant. Learning therefore implies a very different optimal temperature trajectory depending on how the temperature variable is modeled.

The ability of policy to precisely steer the state variables in each period determines the time for which policy sticks to the boundary of the safe region ("state stickiness"). We demonstrate this qualitative difference in policy paths by implementing the climate feedback tipping point in a model without temperature inertia.[19] Figure 5b shows the qualitative difference in the optimal expected temperature curve between this model without inertia and our original DICE extension with inertia. Optimal policy in the model without warming delay keeps temperature constant over most of the next century, whereas optimal policy in our full model shows no such stickiness. In the flat temperature region of the model without inertia, the next unit of emissions causes enough damage (via tipping point risk) to make abatement cost-effective, but the last unit of emission causes far less damage (Figure 5c).[20] Eventually abatement cost falls far enough that it becomes worth abating additional emissions, at which point the damage curves for the next and last unit of emissions merge again. In the full model with inertia, emissions only have a small immediate effect on temperature (reducing the MHE) and optimal policy stabilizes temperatures only once carbon concentrations are already falling. At this point, temperature would remain constant only if emissions increased, which usually is not optimal (Figure 5d). At the same time, the MHE that gives rise to the

---

[19]We model the 5°C climate sensitivity feedback scenario. Temperature is no longer a state variable. We calibrate the model to the same DICE baseline without tipping points by introducing time-dependent feedback processes. Because eliminating warming delay increases marginal damage, we calibrate these feedback processes to suppress the total effect of CO_2 on temperature and so induce policy in line with DICE. The different temperature dynamics and model calibration warn against comparing precise numerical effects between models with and without inertia.

[20]We calculate marginal abatement costs from the optimal abatement policy, and we calculate marginal damage from the value function as in Section 3. We can therefore calculate the value of additional abatement even after reaching full abatement. The differences in these calculations lead to the small offset between the solid line and the dashed and dotted lines over the period when temperature is non-constant. The marginal damage curve jumps down upon eliminating the MHE by keeping temperature constant.
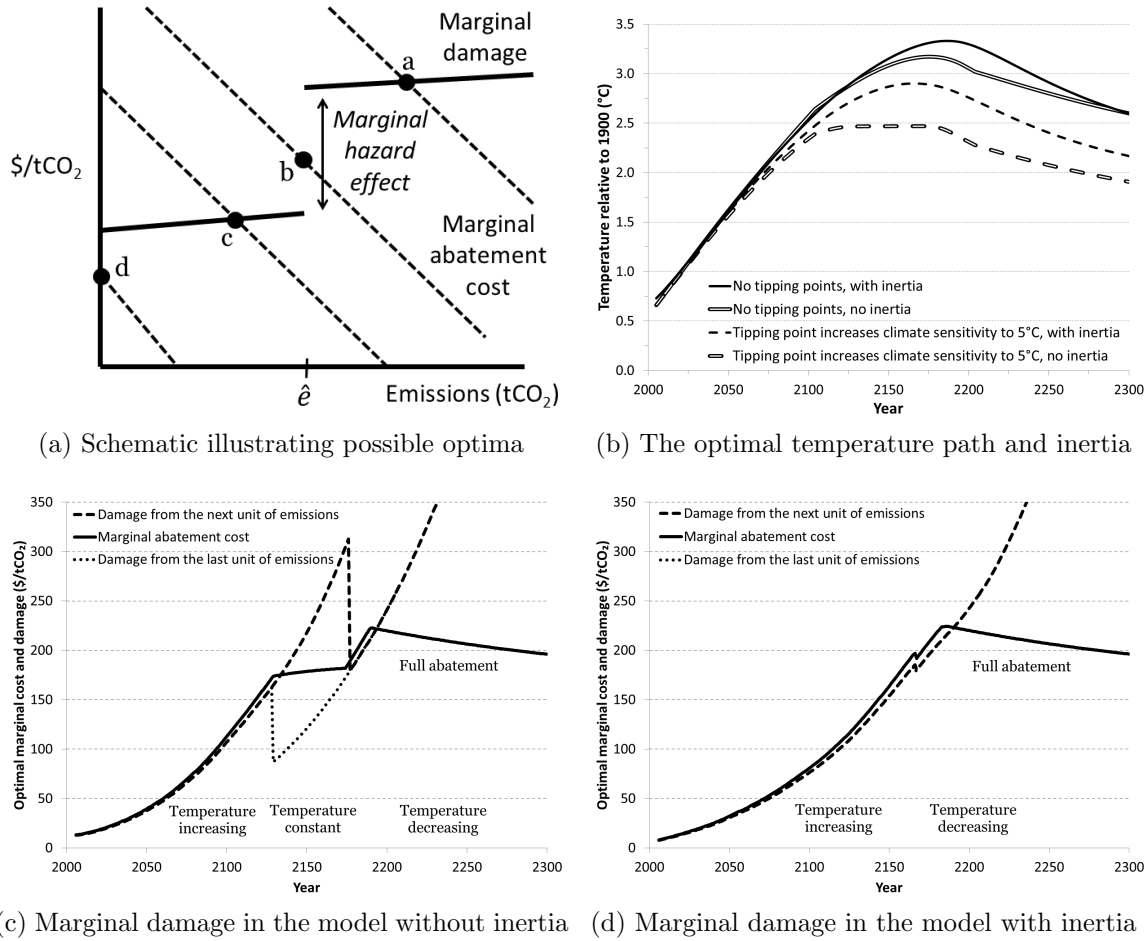
(a) Schematic illustrating possible optima

(b) The optimal temperature path and inertia

(c) Marginal damage in the model without inertia

(d) Marginal damage in the model with inertia

Figure 5: The marginal hazard effect introduces a discontinuity into the marginal damage (marginal benefit of abatement) curve at the emission level $\hat{e}$ that keeps temperature constant (plot 5a). The emission level $\hat{e}$ depends on temperature, CO_2, and time. Optimal emissions are initially great enough to increase temperature (a). As marginal abatement cost (dashed lines) falls over time, temperature becomes first constant (b) and then decreases (c and d). Plots 5b through 5d show how the optimal temperature and marginal damage paths depend on the presence of inertia in the climate system. These plots should be read for the qualitative differences in policy paths, not for quantitative differences that are sensitive to the different model calibrations. The dashed and dotted lines completely overlap in the full model with inertia (plot 5d).

marginal damage discontinuity is smaller because emissions have a smaller impact on the immediate probability of tipping the system.[21]

7   Conclusions

We have shown how to model economic decisions in the face of irreversible tipping points triggered by policy-dependent thresholds. We analytically demonstrated that tipping points affect optimal policy via two channels: the differential welfare impact (DWI) recognizes that today's policy choices also affect welfare in case a tipping point has occurred, and the marginal hazard effect (MHE) recognizes that today's policy choices affect the probability of crossing the threshold. Ambiguity aversion has an ambiguous effect on both terms. In particular, ambiguity aversion amplifies the MHE for small hazards, but can reduce the MHE for large hazards.

Our numerical application developed a dynamic climate-economy model that includes the endogenous possibility of climatic tipping points, endogenous learning about the temperature threshold triggering tipping points, endogenous welfare implications of tipping points, and a generalized welfare evaluation that allows the policymaker to display ambiguity aversion. We find that the possibility of tipping points in the climate system raises the optimal carbon tax. Because of the small annual probability of crossing a climate threshold, these tipping points primarily affect the optimal carbon tax via the MHE, which is amplified by ambiguity aversion. The tipping point increment to the social cost of carbon is not merely a function of the current period's DWI and MHE but is also determined by how current emissions change "tipping lotteries" in all future periods. The climate system's warming delay can trigger future threshold crossings even if future CO_2 concentrations are stable.

Quantitatively, our base case tipping point possibilities can increase the near-term optimal carbon tax by around 40%. The precise effect is sensitive to the type of tipping point, to the strength of the tipping point, and to the distribution for the threshold that triggers the tipping point. Carbon sink tipping points more strongly affect the near-term carbon tax, but climate feedback tipping points have an increasing effect over time and thereby more strongly reduce optimal peak temperature and CO_2. This result demonstrates the value of explicitly modeling tipping points' effects on system dynamics (endogenizing the welfare change). The importance of the MHE demonstrates the importance of endogenizing tipping point possibilities. Ambiguity aversion raises the optimal carbon tax and does so by an

---

[21]In the full model with inertia, additional emissions do affect the probability of crossing the threshold in future periods, but MHE only captures the effect of additional emissions on the probability of crossing a threshold in the current period. The impact of future tipping risk on the social cost of carbon is captured directly in the continuation value.

increasing amount over time.

In a model with endogenous learning, tipping point possibilities create a discontinuity in the future benefits or costs deriving from a control that affects the hazard. This discontinuity is created by the MHE jumping discontinuously from zero to a finite value whenever exploring new, threshold-relevant regions in the state space. It can lead policy to keep the system at the boundary of the safe region for long intervals: going further creates the cost of a possible tipping point. However, if dynamics are governed by a delay equation, inertia can make it inefficient to steer along the boundary. First, inertia in a threshold-relevant state variable limits the ability to steer the state variables along the boundary (making it prohibitively expensive or even impossible). Second, inertia generally reduces the control's impact on the next period's state variables and, thus, reduces the MHE that incentivizes steering exactly along the boundary in the first place. In general, these results demonstrate the value of modeling the effect of learning on optimal policies. In our climate application, we have shown that opting not to model temperature as a delayed state variable produces qualitatively different policy paths. When temperature delay is significant, optimal carbon taxes and temperature do not stay constant for an extended period upon eliminating the hazard.

Our numerical conclusions have implications for economic modeling, climate science, and climate policy. First, economic models of climate change typically assume smooth changes in the climate system. More broadly, nearly all economic models dealing with growth and long-run dynamics assume smoothly changing systems. Those models allowing for discontinuous changes usually incorporate exogenous penalties. We have demonstrated the value of explicitly modeling the shifts in dynamics, showing how feedback and carbon sink tipping points affect optimal abatement policies in different ways. Further, we have shown that the main effect of tipping points on optimal policy is often due to their endogeneity. It is important to model a tipping point's structural effects rather than reducing it to a predetermined shock to utility, and it is important to capture the effect of policy decisions on the tipping point hazard.

Second, our work is a call to the climate sciences to improve knowledge about both the effects of tipping points on system dynamics and the types of temperature paths that trigger them. We have shown that different anticipated changes in dynamics can have quite different effects on the optimal carbon tax. We demonstrate the economic value of scientific information about tipping points and open the door to more comprehensive integrated assessments of abrupt climate transitions. Our sensitivity analysis has shown that these more comprehensive climate change assessments will benefit greatly from progress in the climate sciences that constrains the regions and probabilities of tipping point occurrence.

Third, our findings support the widespread supposition that the existence of tipping points in the climate system should have a strong influence on our current policy decisions. This influence would become stronger if we allowed the policymaker to face multiple tipping points at once. Numerical integrated assessment models are the main quantitative input into regulations that put a price on carbon emissions. Yet past studies omitted climatic features we have shown to be highly relevant. We provide a quantitative basis for adjusting policy for the possibility of tipping points. Much work remains to make tipping representations more realistic, but we have demonstrated how to fully endogenize tipping point possibilities and have provided a first assessment of their effect on policy.

Appendix

A   Additional numerical results for possible climate tipping points

Figure 6 gives additional results for optimal policy in the face of a possible tipping point. As indicated by the main text's results for the optimal tax, the tipping point possibility increases optimal abatement, and abatement jumps down upon first eliminating threshold risk by preventing temperature from increasing. Even though emissions jump up at this point, temperature inertia implies that these greater emissions still decrease temperature. Possible thresholds decrease the optimal investment rate very slightly, but they also eventually lead to slightly greater available output due to reduced climate damages. Threshold possibilities primarily affect the optimal tax via the interaction between current emissions and future threshold crossings. The current period's marginal hazard effect (MHE) is the next largest component, but the current period's differential welfare impact (DWI) is insignificant.

Learning enters the model by expanding the set of safe temperatures and concentrating probability mass on temperatures yet to be explored (Figure 7a). Therefore, as the world reaches higher temperatures, a contemplated temperature increase poses a greater hazard because it cuts through more probability mass (Figure 7b). Lower temperature profiles imply both less learning and flatter hazard functions for each temperature reached (Figure 7c). Optimal policy in the face of possible tipping points lowers the hazard rate by reducing temperature change over time (Figure 7d).[22]

Figure 8 illustrates the policy response to crossing a threshold. It assumes that the policymaker is ignorant of the tipping point possibility until 2075, when the tipping point

---

[22]Each hazard path has a kink at the year 2100 when the exogenous non-CO_2 forcing ceases to increase.

(a) Abatement rate (μ_t)

(b) Abatement cost

(c) Emissions not abated (e_t)

(d) Investment rate

(e) Climate damages (D_t)

(f) Components of the optimal tax
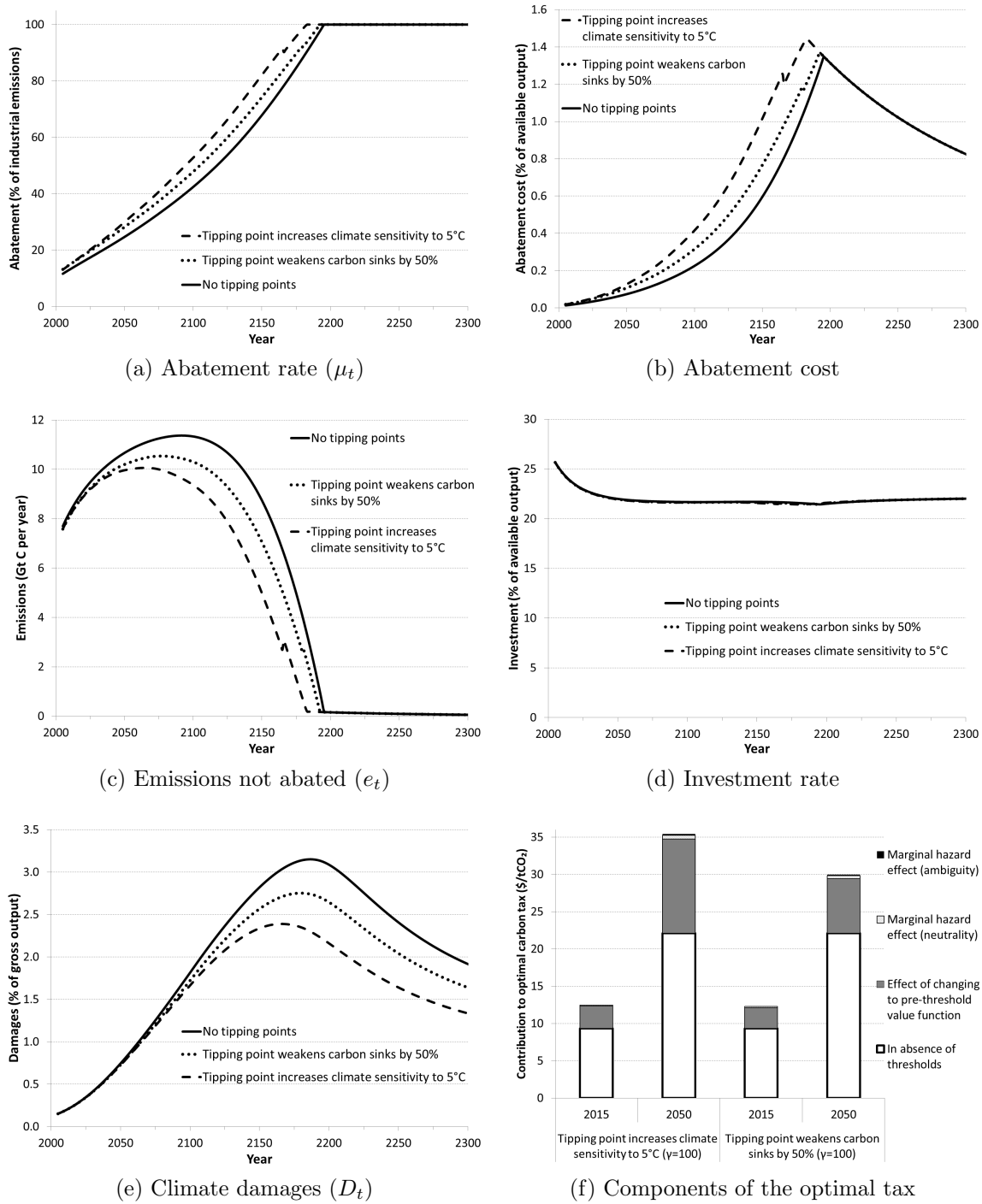
Figure 6: Additional results along optimal policy paths for the base case threshold distribution. The plotted simulations assume expected draws of the temperature shock and also assume that the tipping point never occurs. The lines in the investment and output plots almost completely overlap, and the DWI is too small to plot in the optimal tax decomposition. The decomposition plot includes ambiguity aversion ($\gamma = 100$), and the others use the ambiguity-neutral welfare evaluation.

(a) Base case probability distribution for $\tilde{T}$

(b) Hazard as a function of temperature change

(c) Learning along the optimal path

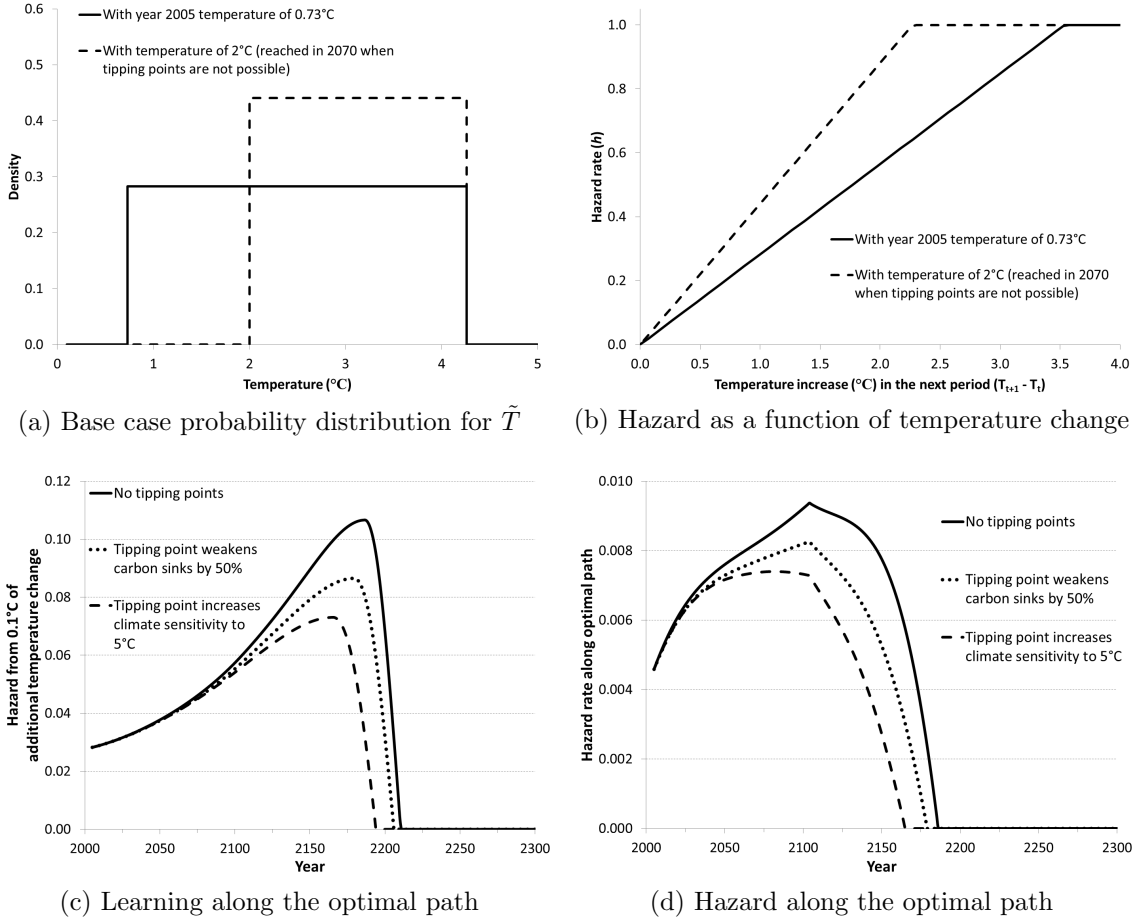(d) Hazard along the optimal path

Figure 7: As the time t expected temperature increases without crossing a threshold, the probability distribution for the threshold level $\tilde{T}$ places more mass on temperatures yet to be reached (a). Each additional increase in temperature therefore also produces a greater risk of crossing the threshold (b). Learning increases the hazard posed by a fixed increment of temperature change as long as temperatures are increasing (c). The actual hazard along the optimal path (d) also depends on the chosen emission policy. The plotted simulations assume expected draws of the temperature shock and also assume that the tipping point never occurs. The hazard in the cases without tipping points is calculated as if they were in fact possible but the decision-maker is unaware of the tipping possibility.

(a) Carbon tax

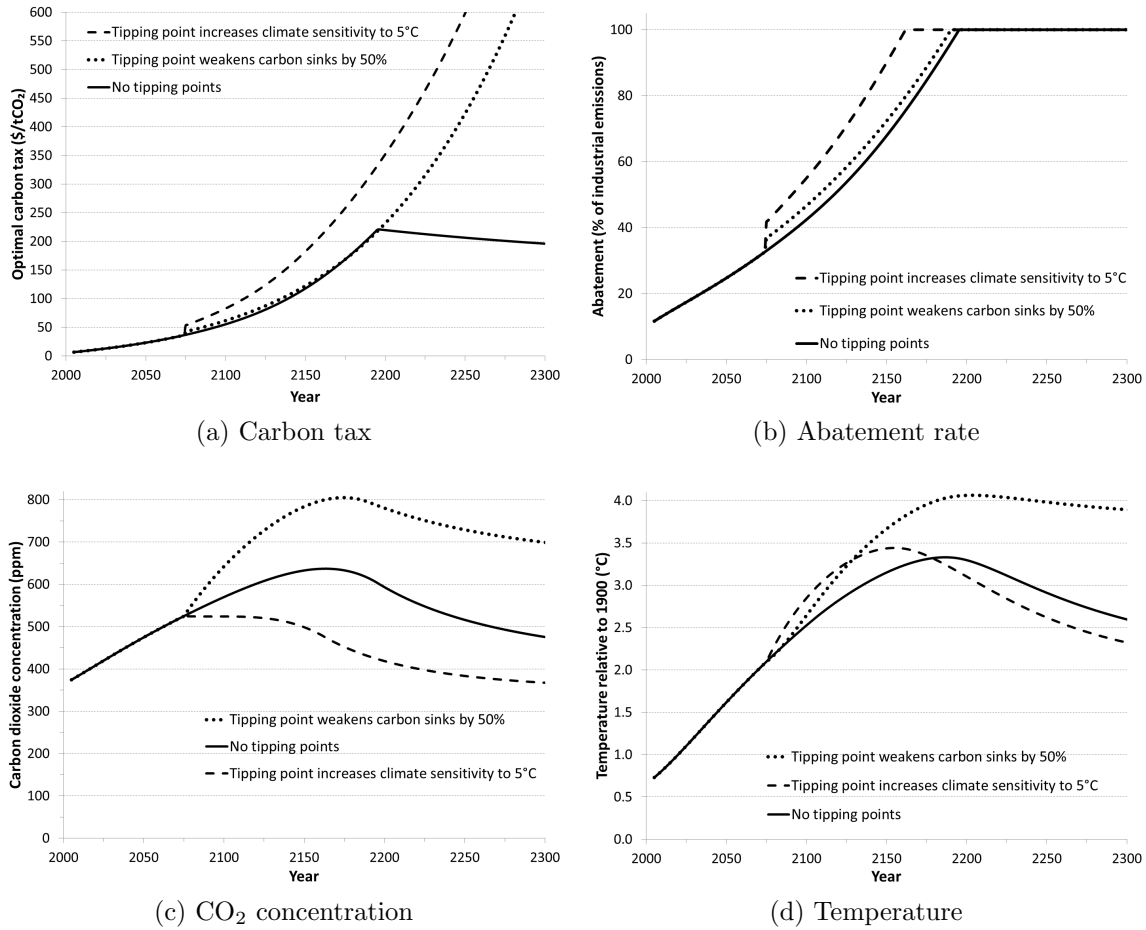(b) Abatement rate

(c) CO_2 concentration

(d) Temperature

Figure 8: The effect of tipping points on post-threshold policy and state variables. Simulations assume the decision-maker is ignorant of threshold possibilities until 2075, when the specified tipping point occurs. The plotted simulations assume expected draws of the temperature shock.

actually occurs. The effect of crossing a threshold depends on how it affects system dynamics and on how policy can compensate for the altered dynamics. The carbon tax responds more strongly to the threshold crossing when it increases climate feedbacks. The CO_2 concentration in the regime with strengthened climate feedbacks follows a lower path than in a case without thresholds, but degraded carbon sinks produce a much higher carbon concentration by increasing the persistence of CO_2 in the atmosphere. Temperature increases under either tipping point, but the more aggressive abatement under the climate feedback tipping point keeps the peak temperature below levels reached under the carbon sink tipping point.

28

B   The dynamic climate-economy model

This appendix provides the complete equations for the climate-economy model extending DICE-2007 from Nordhaus (2008). Table 1 provides the numerical parameterization. The pre-threshold value function is:[23]

$$V_{\psi=0}(k_t, M_t, T_t, t)$$
$$= \max_{c_t, \mu_t} \frac{c_t^{1-\eta}}{1-\eta} + \frac{\beta_t}{1-\eta} \int \left[[1 - h(T_t, T_{t+1})] \left[(1-\eta) V_{\psi=0}(k_{t+1}, M_{t+1}, T_{t+1}, t+1) \right]^{\frac{1-\gamma}{1-\eta}} \right.$$
$$\left. + h(T_t, T_{t+1}) \left[(1-\eta) V_{\psi=1}(k_{t+1}, M_{t+1}, T_{t+1}, t+1) \right]^{\frac{1-\gamma}{1-\eta}} \right]^{\frac{1-\eta}{1-\gamma}} d\mathbb{P}$$

subject to

$$k_{t+1} = e^{-(g_{L,t}+g_{A,t})} \left[(1-\delta_k)k_t + (1 - \Psi_t \mu_t^{a_2}) \frac{Y_t}{1+D_t} - c_t \right] \qquad \text{(Effective capital)}$$

$$M_{t+1} = e_t + M_t \left[\mathbf{b_{11}} + b_{21} \left[\mathbf{b_{12}} + (b_{22} + b_{32}b_{23})\alpha_B(M_t, t) + b_{32}b_{33}\alpha_O(M_t, t) \right] \right] \qquad \text{(CO}_2\text{)}$$

$$T_{t+1} = T_t + C_T \left[F(M_{t+1}, t+1) - \frac{f}{\mathbf{s}} T_t - [1 - \alpha_T(T_t, t)] C_O T_t \right] \qquad \text{(Temperature)}$$

$$c_t + \Psi_t \mu_t^{a_2} \frac{Y_t}{1+D_t} \leq \frac{Y_t}{1+D_t} \qquad \text{(Output constraint)}$$

$$\mu_t \leq 1 \qquad \text{(Non-negativity constraint for emissions)}$$

The state variables are effective capital k_t, atmospheric CO_2 M_t, cumulative temperature change T_t, and time t.[24] Tipping points change the bold parameters. The controls are consumption c_t, abatement μ_t, and, as a residual, investment. Welfare in a given period is the sum of immediate utility $u(c_t) = c_t^{1-\eta}/(1-\eta)$ and the discounted expectation of future welfare. The parameter η is the Arrow-Pratt measure of relative risk aversion, and η^{-1} gives the intertemporal elasticity of substitution. The parameter γ measures aversion to the less confidently known tipping point uncertainty, as described in Section 4. The constraints prevent the decision-maker from using more than the output available after accounting for damages and from abating more than 100% of emissions in a period. When the output constraint is slack, we have positive capital investment, and when the abatement constraint is slack, economic activity produces some CO_2 emissions that are not abated.

---

[23]Normalizing the discount factor, the value function, and some of the equations of motion makes the objective equivalent to maximizing population-weighted per capita consumption $L_t \, u(C_t/L_t)$ as in DICE. The discount factor β_t captures a rate of pure time preference of 1.5% as in DICE-2007. It also adjusts for population growth and technological progress, a step that is part of the normalization yielding a measurement of consumption in effective labor units. See Crost and Traeger (2010) for details.

[24]We solve the model using a transformation mapping the infinite time horizon to the unit interval.

Capital depreciates at rate δ_k, and capital investment comes from any available output not allocated to the control variables of consumption c_t and abatement μ_t. The exogenous variable Ψ_t and parameter a_2 determine the cost of abating the chosen fraction μ_t of emissions. The term outside the brackets in the capital transition equation adjusts for the growth of labor and technology to keep capital in effective terms. Gross output Y_t is a function of the capital stock:

$$Y_t = k_t^{\kappa} \ . \qquad\qquad \text{(Gross output)}$$

The parameter κ gives the capital elasticity in a Cobb-Douglas production function. Climate damages D_t reduce gross output in accord with the total temperature change:

$$D_t = d_1 \left(\epsilon_t T_t\right)^{d_2} \ , \qquad\qquad \text{(Damages)}$$

where the independent, normally distributed multiplicative shock ϵ_t has probability measure $\mathbb{P}$. Optimal policy adjusts current controls in anticipation of possible future shocks, and a given period's realized shock affects the residual output allocated to investment. We calibrate the mean-1 shock to the years 1881-2010 in the NASA Goddard Institute for Space Studies (GISS) Surface Temperature Analysis dataset.[25] We take expected temperature in each year to be the mean of the surrounding 10 years' realized temperatures. The realized standard deviation of the resulting time series of multiplicative shocks is 0.0068.[26] This multiplicative noise captures period-to-period temperature variability that makes extreme outcomes more likely as CO_2 increases.

The carbon dynamics in DICE-2007 are determined by a transition matrix governing the flow between the atmospheric stock (stock 1), the combined biosphere and shallow ocean stock (stock 2), and the deep ocean stock (stock 3). We represent the combined biosphere and shallow ocean stock as a fraction $\alpha_B(M, t)$ of the atmospheric stock and the deep ocean stock as a fraction $\alpha_O(M, t)$ of the atmospheric stock. The parameter b_{12} gives the fraction of atmospheric CO_2 absorbed by land and ocean sinks over a single timestep. The parameter $b_{11} = 1 - b_{12}$ determines the fraction of CO_2 that remains in the atmosphere from period to period, with the remaining terms in the CO_2 transition equation together governing the transfer of carbon from land and ocean sinks back into the atmosphere. As described in Section 4, carbon sink tipping points reduce the parameter b_{12} by a fraction, which also changes b_{11}, α_B, and α_O. For both the pre- and the post-threshold regimes, we run DICE-2007 under several representative emission paths and then approximate the entire coefficient

---

[25]Available at http://data.giss.nasa.gov/gistemp/.

[26]We implement the continuous distribution numerically using a Gauss-Legendre quadrature rule with 8 nodes.

Table 1: Parameterization of the numerical model following DICE-2007. Several values are rounded, and C_T and δ_κ vary slightly over time in order to reproduce the DICE results with an annual timestep.

Parameter	Value	Description
A_0	0.027	Initial production technology
$g_{A,0}$	0.009	Initial annual growth rate of production technology
δ_A	0.001	Annual rate of decline in growth rate of production technology
L_0	6514	Population in 2005 (millions)
L_∞	8600	Asymptotic population (millions)
δ_L	0.035	Annual rate of convergence of population to asymptotic value
σ_0	0.13	Initial emission intensity before emission reductions (GtC/output)
$g_{\sigma,0}$	-0.0073	Initial annual growth rate of emission intensity
δ_σ	0.003	Annual change in growth rate of emission intensity
a_0	1.17	Cost of backstop technology in 2005 ($1000/tC)
a_1	2	Ratio of initial backstop cost to final backstop cost
a_2	2.8	Abatement cost exponent
g_Ψ	-0.005	Annual growth rate of backstop cost
B_0	1.1	Initial non-industrial CO_2 emissions (GtC/y)
g_B	-0.01	Annual growth rate of non-industrial emissions
EF_0	-0.06	Forcing in 2005 from non-CO_2 agents (W m^{-2})
EF_{100}	0.30	Forcing in 2105 from non-CO_2 agents (W m^{-2})
κ	0.3	Capital elasticity in Cobb-Douglas production function
δ_κ	0.06	Annual depreciation rate of capital
d_1	0.0028	Coefficient on temperature in the damage function
d_2	2	Exponent on temperature in the damage function
s	3	Climate sensitivity (°C)
f	3.8	Forcing from doubled CO_2 (W m^{-2})
M_{pre}	596.4	Pre-industrial atmospheric CO_2 (GtC)
C_T	0.03	Translation of forcing into temperature change
C_O	0.3	Translation of surface-ocean temperature gradient into forcing
b_{11},b_{12},b_{13}	0.978,0.023,0	Transfer coefficients for carbon from the atmosphere
b_{21},b_{22},b_{23}	0.011,0.983,0.005	Transfer coefficients for carbon from the combined biosphere and shallow ocean stock
b_{31},b_{32},b_{33}	0,0.0003,0.9997	Transfer coefficients for carbon from the deep ocean
ρ	0.015	Annual rate of pure time preference
η	2	Relative risk aversion (also aversion to intertemporal substitution)
k_0	$137/(A_0 L_0)$	Effective capital in 2005, with 137 US$trillion of capital
M_0	808.9	Atmospheric carbon dioxide (GtC) in 2005
T_0	0.73	Surface temperature (°C) in 2005, relative to 1900

on M_t in the CO_2 transition equation as an interpolated function of CO_2 and time. Time t emissions e_t are given by:

$$e_t = \sigma_t(1 - \mu_t)Y_t + B_t \ . \tag{Emissions}$$

The exogenous variable σ_t is the emission intensity of gross output and B_t gives exogenous CO_2 emissions from non-industrial sources such as land use change.

DICE-2007 determines time t surface temperature from the stock of CO_2, from surface temperature in the previous period, and from the previous period's difference between the surface temperature and the deep ocean temperature. We represent the deep ocean temperature as a fraction $\alpha_T(T, t)$ of surface temperature, where α_T is an interpolated function of temperature and time based on output from DICE-2007 under several representative emission paths. Forcing $F(M_t, t)$ measures the additional energy (W m^{-2}) trapped at the earth's surface by greenhouse gases and other atmospheric agents. Forcing is concave in CO_2:

$$F(M_t, t) = f \ln (M_t/M_{pre}) + EF_t \ , \tag{Forcing}$$

where f is the forcing from doubled CO_2 and EF_t gives the time t exogenous (non-CO_2) forcing. The parameter s in the temperature transition equation is climate sensitivity, or the equilibrium temperature change from doubling CO_2 concentrations. This parameter is altered by climate feedback tipping points as described in Section 4. Finally, the parameter C_O determines how a temperature gradient between the surface and the deep ocean affects forcing at the surface, and the parameter C_T controls the speed with which aggregate forcing changes temperature.

We implement our model with an annual timestep, while DICE-2007 uses a decadal timestep. We therefore adjust all transition equations from those in DICE by calculating the parameter values that would reproduce the state variables' paths if the transition equations were instead applied to an annual timestep with constant policies over the decade. Our model replicates DICE's results when run with DICE's policy path or when optimizing with a 10-year timestep as in DICE. However, when optimizing with an annual timestep and without tipping point possibilities, the policymaker's ability to smooth emissions within a decade leads to the peak CO_2 level being about 30 ppm lower than in DICE, the maximum temperature being about 0.15°C lower, and the year 2015 social cost of carbon being about \$4/tCO_2 lower.

The transition equations for the exogenous variables are as follows. In each case, $t = 0$

corresponds to the year 2005.

$$A_t = A_0 \exp\left[\frac{g_{A,0}}{\delta_A}\left(1 - e^{-t\delta_A}\right)\right] \qquad \text{(Production technology)}$$

$$g_{A,t} = g_{A,0}e^{-t\delta_A} \qquad \text{(Growth rate of production technology)}$$

$$L_t = L_0 + (L_\infty - L_0)\left(1 - e^{-t\delta_L}\right) \qquad \text{(Labor)}$$

$$g_{L,t} = \delta_L \left[\frac{L_\infty}{L_\infty - L_0}e^{t\delta_L} - 1\right]^{-1} \qquad \text{(Growth rate of labor)}$$

$$\beta_t = \exp\left(-\rho + (1-\eta)g_{A,t} + g_{L,t}\right) \qquad \text{(Effective discount factor)}$$

$$\sigma_t = \sigma_0 \exp\left[\frac{g_{\sigma,0}}{\delta_\sigma}\left(1 - e^{-t\delta_\sigma}\right)\right] \qquad \text{(Uncontrolled emissions per output)}$$

$$\Psi_t = \frac{a_0 \sigma_t}{a_2}\left(1 - \frac{1 - e^{tg_\Psi}}{a_1}\right) \qquad \text{(Abatement cost factor)}$$

$$B_t = B_0 e^{tg_B} \qquad \text{(Non-industrial } CO_2 \text{ emissions)}$$

$$EF_t = EF_0 + 0.01(EF_{100} - EF_0)\min\{t, 100\} \qquad \text{(Non-}CO_2 \text{ forcing)}$$

The primary computational challenge in solving the model lies not in finding the optimal actions for a given value function but in determining the value functions that satisfy the relations in equations (1) and (3) (see Kelly and Kolstad, 1999, 2001). We begin with a guess for the value function and a set of Chebychev nodes in the four-dimensional state space. We then use the initial guess for the continuation value to find each node's optimal controls c_t^* and μ_t^* and optimal value. Knowing the optimal value at each Chebychev node, we approximate the value function across the rest of the state space using a set of Chebychev basis polynomials. We repeat the process using this approximated value function as the new initial guess, with iteration continuing until the coefficients of the value approximant's basis functions change by less than 0.0001. We confirm the results by checking that they are robust to variations in the number of nodes and basis polynomials and to variations in the approximated region of the state space.

C   Approximating the effects of ambiguity aversion

We here describe how we analytically approximate the effect of ambiguity aversion on the differential welfare impact (DWI) and the marginal hazard effect (MHE).

First, the contribution of ambiguity aversion to the total DWI was given by:

$$DWI^{ambig} = (1-h)\left(\frac{f'_{amb}(V_{\psi=0})}{f'_{amb}(V_{eff})} - 1\right)\left(-\frac{\partial V_{\psi=0}}{\partial e_t}\right) + h\left(\frac{f'_{amb}(V_{\psi=1})}{f'_{amb}(V_{eff})} - 1\right)\left(-\frac{\partial V_{\psi=1}}{\partial e_t}\right) \quad .$$

Ambiguity aversion increases the DWI when $DWI^{ambig} > 0$ and decreases it when $DWI^{ambig} < 0$. Substituting first-order expansions of f'_{amb} around V_{eff} yields:

$$DWI^{ambig} \approx h \left. \frac{f''_{amb}}{f'_{amb}} \right|_{V_{eff}} \left((V_{\psi=0} - V_{eff}) \frac{\partial V_{\psi=0}}{\partial e_t} - (V_{\psi=1} - V_{eff}) \frac{\partial V_{\psi=1}}{\partial e_t} \right)$$
$$+ \left. \frac{f''_{amb}}{f'_{amb}} \right|_{V_{eff}} (V_{\psi=0} - V_{eff}) \left(-\frac{\partial V_{\psi=0}}{\partial e_t} \right) \quad .$$

We derive an expression for V_{eff} by undertaking a second-order expansion of V_{eff} around $f_{amb}[V_{\psi=0}]$. We then substitute in a second-order expansion of $f_{amb}[V_{\psi=1}]$ around $V_{\psi=0}$ and drop terms of order greater than h to obtain:

$$V_{eff} \approx V_{\psi=0} - h[V_{\psi=0} - V_{\psi=1}] - \frac{1}{2}h \left. \frac{-f''_{amb}}{f'_{amb}} \right|_{V_{\psi=0}} [V_{\psi=0} - V_{\psi=1}]^2 + O(h) \quad . \qquad (11)$$

Substituting for V_{eff} in DWI^{ambig} and again dropping terms of order greater than h yields the result in the text.

Ambiguity aversion changes the MHE through the multiplier

$$\frac{f_{amb}[V_{\psi=0}] - f_{amb}[V_{\psi=1}]}{[V_{\psi=0} - V_{\psi=1}] \, f'_{amb}[V_{eff}]} \quad .$$

Ambiguity aversion increases the MHE when this multiplier is greater than 1. We approximate the multiplier by first approximating f_{amb} by a second-order Taylor expansion around V_{eff}:

$$\frac{f_{amb}[V_{\psi=0}] - f_{amb}[V_{\psi=1}]}{[V_{\psi=0} - V_{\psi=1}] \, f'_{amb}[V_{eff}]} \approx 1 - \left. \frac{-f''_{amb}}{f'_{amb}} \right|_{V_{eff}} \frac{(V_{\psi=0} - V_{eff}) + (V_{\psi=1} - V_{eff})}{2} \quad .$$

Substituting for V_{eff} using equation (11) yields the result in the text.

References

Alley, R. B., J. Marotzke, W. D. Nordhaus, J. T. Overpeck, D. M. Peteet, R. A. Pielke, R. T. Pierrehumbert, P. B. Rhines, T. F. Stocker, L. D. Talley, and J. M. Wallace (2003, March). Abrupt climate change. *Science 299*(5615), 2005–2010.

Archer, D. (2007, July). Methane hydrate stability and anthropogenic climate change. *Biogeosciences 4*(4), 521–544.

Azariadis, C. and A. Drazen (1990, May). Threshold externalities in economic development. *The Quarterly Journal of Economics 105*(2), 501–526.

Brock, W. and D. Starrett (2003, December). Managing systems with non-convex positive feedback. *Environmental and Resource Economics 26*(4), 575–602.

Brozović, N. and W. Schlenker (2011). Optimal management of an ecosystem with an unknown threshold. *Ecological Economics 70*(4), 627–640.

Camerer, C. and M. Weber (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty 5*(4), 325–370.

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

Crost, B. and C. P. Traeger (2010). Risk and aversion in the integrated assessment of climate change. CUDARE Working Paper 1104, University of California, Berkeley.

Davig, T. and E. M. Leeper (2007, June). Generalizing the Taylor principle. *American Economic Review 97*(3), 607–635.

de Zeeuw, A. and A. Zemel (2011, June). Regime shifts and uncertainty in pollution control. Working paper.

Eglin, T., P. Ciais, S. L. Piao, P. Barre, V. Bellassen, P. Cadule, C. Chenu, T. Gasser, C. Koven, M. Reichstein, and P. Smith (2010). Historical and future perspectives of global soil carbon response to climate and land-use changes. *Tellus B 62*(5), 700–718.

Ellison, G. and D. Fudenberg (2003). Knife-edge or plateau: When do market models tip? *Quarterly Journal of Economics 2003*(4), 1249–78.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics 75*(4), 643–669.

Gjerde, J., S. Grepperud, and S. Kverndokk (1999, August). Optimal climate policy under the possibility of a catastrophe. *Resource and Energy Economics 21*(3-4), 289–317.

Greenstone, M., E. Kopits, and A. Wolverton (2011, March). Estimating the social cost of carbon for use in U.S. federal rulemakings: a summary and interpretation. *National Bureau of Economic Research Working Paper Series No. 16913*.

Guo, X., J. Miao, and E. Morellec (2005, May). Irreversible investment with regime shifts. *Journal of Economic Theory 122*(1), 37–59.

Hall, D. C. and R. J. Behl (2006, May). Integrating economic analysis and the science of climate instability. *Ecological Economics 57*(3), 442–465.

Hansen, J., M. Sato, P. Kharecha, D. Beerling, R. Berner, V. Masson-Delmotte, M. Pagani, M. Raymo, D. L. Royer, and J. C. Zachos (2008). Target atmospheric CO2: where should humanity aim? *The Open Atmospheric Science Journal 2*, 217–231.

Heal, G. (1984). Interactions between economy and climate: A framework for policy design under uncertainty. In V. K. Smith and A. D. White (Eds.), *Advances in Applied Microeconomics*, Volume 3, pp. 151–168. Greenwich, CT: JAI Press.

Huntingford, C., R. A. Fisher, L. Mercado, B. B. Booth, S. Sitch, P. P. Harris, P. M. Cox, C. D. Jones, R. A. Betts, Y. Malhi, G. R. Harris, M. Collins, and P. Moorcroft (2008, May). Towards quantifying uncertainty in predictions of amazon 'dieback'. *Philosophical Transactions of the Royal Society B: Biological Sciences 363*(1498), 1857 –1864.

Katz, M. L. and C. Shapiro (1994, April). Systems competition and network effects. *The Journal of Economic Perspectives 8*(2), 93–115.

Keller, K., B. M. Bolker, and D. F. Bradford (2004, July). Uncertain climate thresholds and optimal economic growth. *Journal of Environmental Economics and Management 48*(1), 723–741.

Kelly, D. L. and C. D. Kolstad (1999, February). Bayesian learning, growth, and pollution. *Journal of Economic Dynamics and Control 23*(4), 491–518.

Kelly, D. L. and C. D. Kolstad (2001, October). Solving infinite horizon growth models with an environmental sector. *Computational Economics 18*(2), 217–231.

Keynes, J. M. (1921). *A Treatise on Probability.* London: Macmillan and Co.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005, November). A smooth model of decision making under ambiguity. *Econometrica 73*(6), 1849–1892.

Klibanoff, P., M. Marinacci, and S. Mukerji (2009, May). Recursive smooth ambiguity preferences. *Journal of Economic Theory 144*(3), 930–976.

Knight, F. H. (1921). *Risk, Uncertainty, and Profit.* Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Co.

Kriegler, E., J. W. Hall, H. Held, R. Dawson, and H. J. Schellnhuber (2009, March). Imprecise probability assessment of tipping points in the climate system. *Proceedings of the National Academy of Sciences 106*(13), 5041–5046.

Lange, A. and N. Treich (2008, July). Uncertainty, learning and ambiguity in economic mod-

els on climate policy: some classical results and new directions. *Climatic Change 89*(1), 7–21.

Le Quéré, C., C. Rodenbeck, E. T. Buitenhuis, T. J. Conway, R. Langenfelds, A. Gomez, C. Labuschagne, M. Ramonet, T. Nakazawa, N. Metzl, N. Gillett, and M. Heimann (2007, June). Saturation of the southern ocean CO2 sink due to recent climate change. *Science 316*(5832), 1735–1738.

Leach, A. J. (2007, May). The climate change learning curve. *Journal of Economic Dynamics and Control 31*(5), 1728–1752.

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

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

Millner, A., S. Dietz, and G. Heal (2010, June). Ambiguity and climate policy. *National Bureau of Economic Research Working Paper Series No. 16050*.

Miranda, M. J. and P. L. Fackler (2002). *Applied Computational Economics and Finance*. Cambridge, Massachusetts: MIT Press.

Nævdal, E. (2006, July). Dynamic optimisation in the presence of threshold effects when the location of the threshold is uncertain—with an application to a possible disintegration of the western antarctic ice sheet. *Journal of Economic Dynamics and Control 30*(7), 1131–1158.

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

Nordhaus, W. D. (2008). *A Question of Balance: Weighing the Options on Global Warming Policies*. New Haven: Yale University Press.

Overpeck, J. T. and J. E. Cole (2006). Abrupt change in earth's climate system. *Annual Review of Environment and Resources 31*(1), 1–31.

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

Ramanathan, V. and Y. Feng (2008). On avoiding dangerous anthropogenic interference with the climate system: Formidable challenges ahead. *Proceedings of the National Academy*

of Sciences 105 (38), 14245 –14250.

Schaefer, K., T. Zhang, L. Bruhwiler, and A. P. Barrett (2011, April). Amount and timing of permafrost carbon release in response to climate warming. *Tellus B 63* (2), 165–180.

Schelling, T. (1971, July). Dynamic models of segregation. *The Journal of Mathematical Sociology 1* (2), 143–186.

Skiba, A. K. (1978, May). Optimal growth with a convex-concave production function. *Econometrica 46* (3), 527–539.

Smith, J. B., S. H. Schneider, M. Oppenheimer, G. W. Yohe, W. Hare, M. D. Mastrandrea, A. Patwardhan, I. Burton, J. Corfee-Morlot, C. H. D. Magadza, H. Fussel, A. B. Pittock, A. Rahman, A. Suarez, and J. van Ypersele (2009). Assessing dangerous climate change through an update of the intergovernmental panel on climate change (IPCC) "reasons for concern". *Proceedings of the National Academy of Sciences 106* (11), 4133–4137.

Traeger, C. P. (2010). Subjective risk, confidence, and ambiguity. CUDARE Working Paper 1103, University of California, Berkeley.

Tsur, Y. and A. Zemel (1996). Accounting for global warming risks: Resource management under event uncertainty. *Journal of Economic Dynamics and Control 20* (6-7), 1289–1305.

Wagener, F. O. O. (2003, July). Skiba points and heteroclinic bifurcations, with applications to the shallow lake system. *Journal of Economic Dynamics and Control 27* (9), 1533–1561.